DOCUMENT RESUME

ED 163 981 IR 006 660

AUTHOR	Braunstein, Yale M.; And Others
TITLE	Economics of Property Rights as Applied to Computer Software and Data Bases.
INSTITUTION	New York Univ., N.Y.
SPONS AGENCY	National Commission on New Technological Uses of Copyrighted Works, Washington, D.C.
REPORT NO	CONTU-77-0005
PUB DATE	Jun 77
NOTE	131p.; For related documents, see IR 006 669-671; Some pages may not reproduce well due to quality of type in original
EDRS PRICE	MF-$0.83 HC-$7.35 Plus Postage.
DESCRIPTORS	*Computer Programs; *Copyrights; Cost Effectiveness; Costs; *Data Bases; *Economic Factors; Electronic Data Processing; Federal Legislation; Futures (of Society); Information Systems; Mathematical Models
IDENTIFIERS	Commission on New Technological Uses; *Computer Software

ABSTRACT

Four major discussions are presented related to economic issues involved with property rights as applied to computer software and data bases. The first discussion qualitatively compares the relative incentives to computer software producers found in alternative regimes of copyright and trade secrecy. The latter is found to have certain dysfunctional qualities which copyright might ameliorate. The second discussion providing a critical view of exemptions from copyright for broad classes of users arrives at the conclusion that exemptions decrease economic efficiency which would be increased through user charges. The third section creates a model of software production and protection and indicates that the term of copyright protection for software should be very short. The final discussion analyzes the trade-off between the scope of protection and the length of time for which the protection lasts. Appendices include mathematical derivations of the optimal protection period for a linear demand curve and a nonlinear demand curve. (Author/CMV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163981

PB 268 787

74

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ECONOMICS OF PROPERTY RIGHTS AS APPLIED TO COMPUTER SOFTWARE AND DATA BASES

by

Yale M. Braunstein
Dietrich M. Fischer
Janusz A. Ordover
William J. Baumol

New York University
(Revised, June 1977)

IR 006 660

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

Report prepared for the National Commission on New Technological Uses
of Copyrighted Works (CONTU)

BIBLIOGRAPHIC DATA SHEET

1. Report No. CONTU 77-0005

2.

3. Recipient's Accession No.

4. Title and Subtitle
Economics of Property Rights as Applied to Computer Software and Data Bases (Report prepared for the National Commission on New Technological Uses of Copyrighted Works)

5. Report Date
June 1977

6.

7. Author(s) Braunstein, Yale M.; Fischer, Dietrich M.; Ordover, Janusz A.; and Baumol, William J.

8. Performing Organization Rept. No.

9. Performing Organization Name and Address

New York University

10. Project/Task/Work Unit No.
Pub. L. 93-573

11. Contract/Grant No.

12. Sponsoring Organization Name and Address
National Commission on New Technological Uses of
Copyrighted Works (CONTU)
Washington, D.C. 20558

13. Type of Report & Period Covered
Economic analysis

14.

15. Supplementary Notes

16. Abstracts

The report includes four major discussions the first of which qualitatively compares the relative incentives to computer software producers found in alternative regimes of copyright and trade secrecy. The latter is found to have certain dysfunctional qualities which copyright might ameliorate. The second discussion provides a critical view of exemptions from copyright for broad classes of users. The authors hold that exemptions decrease economic efficiency which would be increased through user charges. The third section creates a model of software production and protection and indicates that the term of copyright protection for software should be very short. The final chapter of the piece analyzes the trade-off between the scope of protection and the length of time for which the protection lasts. It finds that the more there is of one the less there needs to be of the other.

17. Key Words and Document Analysis. 17a. Descriptors

Software
Computerized Data Bases
Computers
CONTU
Economics
Pricing
Trade Secrets
Property Rights

Copyright

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement

Release Unlimited

19. Security Class (This Report)
UNCLASSIFIED

20. Security Class (This Page
UNCLASSIFIED

22. Price PCA07
MFA01

SUMMARY

The Major Issues

This report addresses two issues, one general and one specific. The general issue is whether economic analysis can provide insights into the proper scope of copyright coverage for works stored in or used with the assistance of computers or other automated data processing systems. The more specific issue is whether computer software should be copyrightable and, if so, what the terms of that copyright should be.

Advantages of Copyright

We show that, if one wishes to encourage the private provision and use of scientific and artistic works, the net value of output of goods and services to the public as a whole can be increased by the broadest possible specification of property rights by means of copyright. This conclusion remains valid even when one considers the economics of modern computer systems.

Our analysis proceeds by showing how the economics of investment in the product of intellect influences the structure of the market for those works. Economic theory is then used to show that in such a market there is an optimal structure of prices. Here "optimal" means that economic welfare of society is maximized given the requirement that the producers break even or make some minimum level of profits. We then show that this implies that fees for the usage of copyrighted material be charged for as many of the uses as possible. This then leads to the conclusion that if any groups of users are exempted from paying, those users who are charged will have to pay more than they otherwise would (or the profits of the copyright owners and the incentives for the production of new works will be reduced).

There is one possible exception to the general rule that charges should be levied in every market. This exception arises when the costs of collection of user fees are likely to be so high relative to the sums involved as to render that collection uneconomical. These administrative costs are commonly referred to as "transactions costs". In computer information systems, costs of storing, transforming or manipulating data are generally much lower per datum supplied than are those of non-computer systems. Moreover, costs of monitoring the usage of those data also turn out to be correspondingly low. We conclude that the transactions costs are, in general, sufficiently insignificant, so as not to invalidate the argument for user fees and broad copyright protection.

Copyright vs. Trade Secrecy

The two prime methods available for the protection of computer software -- copyright and trade secrecy -- have opposite effects on usage. Whereas copyright promotes disclosure, reliance on the trade secrecy statutes requires suppression of information on advances in software design. We analyze these effects and infer that the use of copyright for protection will have beneficial effects on the quality and variety of software.

If producers have to rely on trade secret laws, they will have an incentive to keep their programs and data bases secret to prevent unauthorized use. Programs will tend to be written in obscure form, which makes it difficult to modify them or transfer them to a different computer. Potential buyers of computer software will lack information on what is available. Competition will lead to a duplication of efforts. Too many resources will be devoted to the development of software which is similar to existing products, so that a firm can capture a portion of a lucrative market, at the expense of others. Too few resources will be devoted to highly innovative and more risky software developments, which are beneficial to society, but less profitable to an individual firm.

Copyright protection has the advantage of giving better market information to software users; it encourages the development of software which can be used for many purposes and on many different computers since that gives a producer a wider market without fear of unauthorized use of his products; it reduces duplication of efforts; and it will tend to lead to the availability of a wider variety of products.

Choice of Copyright Period

For how long a period should a copyright of computer software or data bases be granted? On the one hand, the longer the period of protection the more profitable is the development of new software and the more will be produced. But the temporary monopoly during which a producer can charge a price for the use of his products also discourages some potential users, and in this way reduces the social usefulness of whatever amount of software is produced. The optimal time span for the

C

period of protection yields the best compromise between these two effects. This optimal period depends on a number of economic and technological factors. In this report formulas are given which permit computation of an optimal protection period from empirical data. The following four factors all tend to favor a short protection period: a short useful life-span of software packages; rapidly increasing costs with an increase in the production of software; demand which is very responsive to price changes; and a high discount rate. Of course, the opposite factors all favor a longer protection period.

The actual choice of optimal protection period will require further research in gathering the pertinent data. It is research that is surely worth undertaking because the copyright period selected can make a substantial difference in the general welfare.

Other Conclusions

It should be noted that there is empirical evidence which implies that for the encouragement of the development of truly innovative software and fundamental breakthroughs in software design public support of research institutions will continue to be needed even if copyright protection is provided.

In any event, it is clear that encouragement of software design, whether through copyright protection or by other means, is important as an element in the stimulation of economic growth. Declining costs of computer hardware and expanding opportunities for automation imply that computer hardware and software together will play an increasingly crucial role in our economic future.

TABLE OF CONTENTS

I. OVERVIEW OF THE ISSUES

- A. Historical Perspective
- B. Specific Issues in the Area of Computer-Based Information
- C. Introduction to the Economic Analysis
- D. Issues in Copyright Policy
- E. Copyright and Its Alternatives
- F. Terms of Copyright
- G. Compromises Required in Practice

II. ECONOMICS OF PROPERTY RIGHTS AS APPLIED TO COMPUTER DATA BASES AND SOFTWARE

- A. Economics of Pricing: Some Basic Principles
- B. Costs in the Information Industries
  - 1. Examples from the scientific journal publishing sector
  - 2. Scale economies in computer software and computer data bases
- C. Setting Prices to Cover Costs
- D. The Implications of Quasi-Optimal Pricing for the Statutory Protection of CS and CDB
- E. Some Extensions of the Theory of Quasi-Optimal Pricing and Their Relevance for the Structure of Copyright Laws
- F. Transactions Costs in Computer Information Systems

III. THE CONSEQUENCES OF TRADE SECRECY AND COPYRIGHT PROTECTION FOR THE QUALITY, VARIETY, AND QUANTITY OF NEW PRODUCTS

- A. Qualitative Effects of Trade Secrecy and Copyright Protection
- B. Quantitative Effects of Trade Secrecy and Copyright Protection

IV. THE OPTIMAL DURATION OF COPYRIGHT

- A. Factors Which Influence the Optimal Duration of the Monopoly
- B. A Model to Simulate the Behavior of a Typical Software Firm
- C. Results Obtained from the Model and Their Policy Implications
- D. Data Requirements and Estimates of Optimal Protection Periods
- E. Protection vs. Public Subsidies for the Production of Software

APPENDICES

1. Derivation of the Optimal Protection Period for a Linear Demand Curve
2. Derivation of theOptimal Protection Period for an Example of a Nonlinear Demand Curve

I

OVERVIEW OF THE ISSUES

A. Historical Perspective

Since the passage of the 1909 copyright act (and in some ways, since the enactment of the first U.S. copyright law in 1790) it has been necessary for the courts and the Copyright Office to interpret that act in the light of technological and business advances. Two trends have emerged: First, utilitarian (non-artistic) creations and compilations of data such as interest tables and telephone books have been found to be copyrightable.[1,2] Second, the relationship of the copyright act to the products of and information transmitted by new technologies has frequently created difficult cases and often bad, or at least unclear, law. Among these technological advances have been sound recordings, radio and television broadcasts, photocopying, and cable television.[3]

Although the Supreme Court has expressed "the conviction that encouragement of individual effort by personal gain is the best way to advance public welfare through the talents of authors and inventors..,"[4] the decision concerning new techniques have sometimes appeared to ignore that conviction. (Fortnightly and Williams and Wilkins are perfect examples.)

---

[1] Edwards and Deutch Lithographing Co. v. Boorman, 15 F. 2d 35(7th Cir. 1926).

[2] Leon v. Pacific Tel. & Tel. Co., 91 F. 2d 484 (9th Cir. 1937).

[3] Among the more distinctive cases are: White-Smith Music Publishing Co. v. Apollo Co., 209 U.S. 1 (1908) (Piano roll not a copy of copyrighted music); Capitol Records, Inc., v. Mercury Records Corp., 221 F. 2d 657 (2d Cir. 1955) (phonograph record not copyrightable); Williams & Wilkins Co. v. United States, 420 U.S. 376 (1975) (Wholesale governmental photocopying no infringement); and Fortnightly Corp. v. United Artists Television, Inc., 392 U.S. 390 (1968) (CATV system does not "perform" works it makes available to its customers.)

[4] Mazer v. Stein (347 U.S. 201, 1954).

As a result several copyright revision bills have in recent years been considered by Congress; these addressed various parts of the issue of the proper scope of copyright protection in light of new technology.[1] The 1969 revision bill (S.543) and the version finally enacted in 1976 (P.L. 94-553) each contained a section (§ 117) that stipulated that the case law on the use of copyrighted works in computer systems would be unaffected by the new law. In an action which may have permitted the passage of the new copyright act, the National Commission on New Technological Uses of Copyrighted Works (CONTU) was established to recommend legislation in the area of the reproduction and use of copyrighted works by computer and photocopying systems.[2]

In general terms the conflict to be resolved is that between the desire for a free flow of ideas, on one hand, with the establishment of incentives for the creation of ideas on the other. This balance is addressed by the copyright and patent laws, both of which grant, under certain circumstances, limited monopolies to the creators of inventions, books, and other forms of products of intellectual creativity.

B. Specific Issues in the Area of Computer-based Information

There are three basic questions that underlie this discussion of the proper scope of copyright protection:

(1) What is the product, i.e., what exactly is copyrightable in the context of items used with the aid of a computer system?

(2) How do we define usage in this context?

---

[1] For example, see H.R.2512 and S.597 (1967), S.543 (1969).

[2] P.L.93-573, enacted on Dec. 31, 1974. The Commissioners were appointed on July 25, 1975.

(3) How do we detect and monitor such usage?

These questions obviously do not pertain only to information used in a computer or automatic data processing system. There are natural parallels, say, to the use of copyrighted phonographic recordings by radio stations or the transmission of copyrighted still or motion pictures by means of television. But in the computer field these questions can not simply be answered by analogy with other forms of communication.

Although the Commission has received considerable testimony and advice on whether input to, storage in, or output from a computer whould constitute usage,[1] we shall take a step back to the basic underlying principle which has given rise to these questions:

Principle 1.1 The more complete and certain the specification of property rights, the greater the level of economic efficiency that is possible.

It is because of this principle that one seeks to define what is meant by property in reference to information used by a computer system. As the Data Base Subcommittee has indicated, it seems clear that the intention of Congress was that computer data bases be considered a proper subject for copyright.[2] However in the discussions of proprietary rights in computer programs and software, arguments have been made for protection by means of copyright, patent, or the trade secret laws.[3] It is in the comparison between protection by means of copyright and reliance on trade secrecy and restrictive licensing that the contrasts are most apparent. Copyright promotes disclosure while trade secrecy results in minimal disclosure and dissemination (if it works).

---

[1] See, e.g., Data Base Subcommittee report, Feb. 18, 1977.

[2] *Ibid.*, p.l.

[3] For example, see D. Bender, "Trade Secret Protection of Software," *G. Washington Law Review* 38:909 (1970); C.N. Mooers, "Computer Software and Copyright," *Computing Surveys* 7 (March 1975), 45-72; and O.R. Smoot, "Development of an International System for Legal Protection of Computer Programs," *Communications of the ACM* 19 (April 1976), 171-174.

We shall examine the economic issues such as the trade-off between the incentives to create and the incentives to use new information or products. Although we shall not be able to disregard completely many of the legal issues, they shall not be in the forefront of our discussion.

C. Introduction to the Economic Analysis

The subject of copyright gives rise to a number of complex economic issues. After pointing out some fundamental relationships that will underlie our discussion, it is well to begin with a survey of these issues. From there we turn to a brief indication of the evaluation of these issues to which one is led by economic analysis. Our report then turns to a rather more probing discussion of a number of these matters, and ends up with a careful description of a novel and powerful method by which one can evaluate one of the key issues in the area -- the optimal length of copyright.

The need for copyright or some substitute arises from what economists call difficulty of exclusion from the use of printed material, software, and various other such products whose initial cost of production is quite high but for which the cost of replication is very small. Ability to exclude means that the producer of such an item can prevent a potential user from employing it unless that person is willing to provide payment for the item. Difficulty of exclusion means that people can help themselves to the item without payment and with little or no fear of untoward consequences. Such goods, particularly when they are expensive to produce, normally will find no private suppliers. Firms will surely be discouraged from investing in new software if the products will be immediately available to anyone for the taking. To prevent this is, of course, the purpose of copyright. It is intended to protect the investment of firms that put resources into valuable new products, not only as a means to protect the interests of those firms, but perhaps even more important from the social viewpoint, to encourage the production of valuable items from whose use it would be difficult to exclude anyone without such protection.

However, there is another side to the matter which will play an important role later in the analysis. These same products which seem to require statutory protection are also, generally, items which, in the view of economists, should be offered with maximal encouragement of widespread use. That is, these items generally involve a heavy (sunk) cost of development to which little or nothing is added when there is an increase in number of users. This is in sharp contrast with other products such as food or clothing in which additional usage requires substantial additional use of resources -- additional users are not free. Thus, economists argue, while goods like shirts or potatoes should have a high price reflecting the high costs imposed by additional users, there is something undesirable about a commensurately high price for the use of a software package since that high price will discourage widespread use even though such added use costs society little or nothing over and above the sunk cost of development of the package.

D. Issues in Copyright Policy

Having discussed the logic of the underlying problem we can turn next to a listing of the issues which must be faced by copyright policy. In the following section we will see how the analysis which has just been summarized helps us to provide a fruitful evaluation of these issues.

The most obvious issue that arises in this area is whether copyright is the only way to go about dealing with the problem that copyright protection is designed to solve. As a matter of fact there are several alternatives to copyright, and the issue is to determine whether anyone of them as generally better than the others, or, what may be more likely, whether some work well in some circumstances while others are better suited under different conditions.

Since copyright is intended to protect the interests of those who invest in the design of certain types of new products and to encourage that sort of investment, an obvious alternative to copyright is the provision of added protection to industrial secrets. If a potential investor in a software package can be made to feel reasonably confident that others will be unable to discover the secrets underlying his product so that he will be able to retain a monopoly over its use, then that may perhaps be as effective a stimulus to this sort of investment as copyright protection.

There is also a third possibility where neither secrecy nor copyright protection is applicable or effective, then private investment will obviously be discouraged. In that case the substitute possibility is investment by the public sector. This helps to explain the heavy investment of government agencies in a wide variety of types of research, most notably in basic research which is preponderantly carried out under government sponsorship.

Thus there are alternatives to copyright, and we will see in the next section that for each of them there are areas in which it promises to be more appropriate than the others.

The second set of issues that arises in the area of copyright policy is the nature of the terms that should be permitted. Here several issues arise immediately:

(1) Over how long a period should the copyright extend?

(2) Should the length of copyright be uniform for all items protected?

(3) Should there be any restrictions upon the monopoly power conferred upon the holder of a copyright?

(4) In particular, should there be compulsory licensing of the copyrighted product?

(5) If there is compulsory licensing, how should the price be set?

(6) Should there be classes of uses or classes of users of copyrighted products who are exempted from the copyright provisions?

Finally, there are a variety of issues that arise in carrying out copyright policy in light of the complications inevitably injected by reality. These force policy-makers to accept compromises which can be avoided altogether only in theoretical discussions. Only two examples will be offered, both of them arising because of problems of enforcement.

First, are there types of use which should be exempted from copyright protection, not because their exemption is inherently desirable but because the cost of entering copyright protection in these areas exceeds the potential benefits of enforcement? If so what can one say about the nature of such exemptions? (This is one of the issues addressed by the "fair-use" section (§107) of the 1976 Copyright revision.)

Our second example relates to charges for the use of copyrighted material. Here the question is whether the imposition of charges based directly on use is always practical, and if not, what alternatives constitute a second-best solution.

This brief discussion has been intended to provide a short list of the main economic issues involved in copyright policy. It certainly does not pretend to be exhaustive nor does it even begin to suggest the noneconomic (e.g., legal, political and social) issues. This list should, however, serve as an adequate working framework for our discussion.

We proceed now to a brief commentary upon each of these issues, indicating how economic analysis suggests they should be viewed. After that, we will turn to a somewhat more extensive examination of several of the more crucial topics.

E. Copyright and Its Alternatives

We have indicated that copyright is only one of three avenues to encourage investment in the development of the pertinent new products, trade secrecy and direct government investments being the two prime alternatives. What is the appropriate role of each of these instruments?

Later we will discuss in some detail the differences in the consequences of secrecy and copyrighting. But, in brief summary, secrecy necessarily restricts the range of direct users of the product involved, whereas copyright does not necessarily do so and certainly does not do it where there is compulsory licensing at reasonable prices. Second, resort to secrecy does encourage expenditure on product features which inhibit discovery of the secrets. There are other costs, extending even to outlays on espionage which this process encourages. On the other hand, if the copyright remains valuable as its expiration data approaches, this can encourage research expenditures whose only purpose is to discover product variations just sufficient to justify a new copyright, an exercise which obviously has little payoff to society.

Secrecy is likely to work most effectively where the uncopyrighted product is an intermediate good which is itself used in the production of other goods, e.g., if it is a computer program used to process other people's data. When this is the case it is easier for the investor to earn the cost of his investment without revealing his secret because his product need never leave his hands. On the other hand, if the investor must rely for his earnings on the sale of the uncopyrighted product itself (e.g., by selling the program itself to users) then it is very difficult to retain secrecy. This distinction naturally suggests the sort of item in which innovation can be expected in the absence of copyright protection, and the sort of product which requires copyright if private investment in development is to be obtained.

But there are reasons why the third way of obtaining this sort of investment -- direct government financing — should at least sometimes be utilized. We saw earlier that for at least some potentially copyrightable products the cost of serving additional users is negligible, so that a high price which discourages use is undesirable socially. However a very low price is likely to be incompatible with recovery of initial investment outlays for its development. Where this is a significant consideration (and precise conditions for this case can be enunciated -- see Baumol and Ordover [1977] appended to this report),[1] then neither copyright nor secrecy will be in the social interest, for both operate via the extraction of prices sufficiently high to bring a return that will be attractive to private investment. It is clear that in such a case, if the investment in the product is worthwhile to society but compensatory pricing is not, then the only alternative is government financing. Even if one were to think of another method of making such investment attractive financially to private enterprise, it would, by definition, be no more satisfactory than the use of copyright or reliance on secrecy.

---

[1] Baumol, W.J. and Ordover, J.A., "On the Optimality of Public-Goods Pricing with Exclusion Devices," Kyklos 30 (1977) Fasc. 1, pp. 5-21 (copy attached).

F. Terms of Copyright

We come next to the issue of which conditions should be applied to the granting of a copyright where that is the procedure adopted.

The first issue to be considered is the period of time over which the copyright should remain valid. It is shown in detail in this report that this is an issue which need not be settled *ad hoc*, on the basis of tradition or compromise. There are strong economic considerations which can serve as an underpinning for an analysis of optimal copyright life. The point is that a very short copyright period will fail in its purpose by being insufficient to attract any significant amount of investment into the development process. On the other hand a very long copyright period can extract an excessive price from society by granting monopoly powers to investors for an excessive period. The optimality calculation consists of a balancing of these two considerations to determine that duration of copyright which maximizes its net contribution to social welfare.

This immediately indicates that, at least in principle, the optimal period of protection will vary from case to case. It will depend on the annual profitability of the investment, the magnitude of the monopoly price extracted from the public, the "natural" rate of obsolescence of the product of the investment[1] and a variety of other such considerations which are indicated specifically in detail later in our analysis. That means that one industry or one type of product may call for a copyright life different from another. A computer program may be encouraged most effectively by a copyright lasting x years while a novel may call for a y year copyright. Obviously, a case-by-case approach to copyright period is totally impractical, but it is certainly worth considering whether the establishment of, say, two or three different categories of materials eligible for copyright protection, each with its own copyright period, may not be desirable and feasible administratively. Often it has proven to be true that where administrative simplicity calls for uniformity while other efficiency considerations call for variability, a very small number of categories add very little to administrative costs and yet contribute enormously to efficiency -- sometimes capturing the bulk of the gains that are potentially available. It is because of this possibility that this report devotes so much space to a model for the evaluation of optimal copyright duration. Should it be determined that the possibility of several copyright periods is worth further exploration, our model indicates just how this issue should be investigated.

Other potential provisions of the copyright that require comment are the idea of compulsory licensing, the possibility of restrictions upon pricing and the notion of exemptions from the terms of the copyright of certain categories of use or certain categories of users.

---

[1] In general, for example, the optimal period of protection will be less than the useful economic life of the item covered. This is necessary in order to make sure that the net gains from the project are shared by the investor and by the general public. Obviously, the latter's gains are limited until after the expiration of the copyright since before that time the copyright holder may be able to use the monopoly power it confers to extract the bulk of the gains flowing from the copyrighted product.

All of these possibilities share one common feature. Each of them constitutes some restriction upon the avenues which the holder of the copyright can use to pursue profits. This means that, if statutory protection is to serve as an incentive for investment in the production of material subject to that protection, any provision of the sort now under discussion must weaken this incentive.[1] To achieve any given degree of stimulation of such investment, a price ceiling, a requirement of licensing or other such restrictions must be offset by a commensurate improvement in some other inducement, e.g., it may call for a countervailing increase in length of the period of copyright protection. But we have seen that a lengthening of this period is not without its social cost, so that the decision becomes a trade-off in which the advantages expected to flow from the restrictions in the copyright holder's monopoly power are balanced against the social cost of the lengthened protection period.

Compulsory licensing and restrictions upon the price charged for a license obviously must constitute a gain to those who decide to acquire a license as well as to the customers for their products. This will necessarily be so since those who obtain a license do so voluntarily, and hence to them it must be worth its price. Without compulsory licensing they would be deprived of this gain.

But one must be careful not to jump to the conclusion that this is a net gain to the community. It is equally certain that compulsory licensing must constitute a net loss to the holder of the copyright, at least in his own view of the matter, for otherwise there would be no together rather than separately.) The interrelation between scale economies and production complementarities can be seen in the operations of the abstracting and indexing services. For example, the American Society of Civil Engineers produce printed and machine readable abstracts and indices as part of the same operations. The development costs for this computer-assisted abstracting and indexing system were approximately $25,000. It now costs less than $2,000 per month for the input of 100 items.[1]

---

[1] This argument only considers the economic incentives for investment in the production of intellectual works. There are, of course, other incentives.

The econometric studies reported in this section lead us to the following tentative conclusion: Most STI systems -- scientific journals, CS, CDB -- are characterized by scale economies and in many cases by production complementarities.

C. Setting Prices to Cover the Costs

We have seen already that marginal cost pricing has desirable properties from the standpoint of the society. Thus, if price is equal to marginal cost, all those who are willing to pay the price are per force willing to pay for the resource which is used to provide them with the commodity in question. Ideally, then, we would like to have all the commodities priced in this manner. For a single-product firm, or for a multi-product firm in which all fixed costs can be rationally allocated among various products, i.e., when fixed costs are product specific, the simplistic prescription is to switch from marginal cost pricing to the average cost pricing. By definition, a firm that charges a price equal to or exceeding the average cost can recover all its costs. Setting aside the difficulties associated with the definition of the say, to certain classes of nonprofit users, but there seems to be no reason to require the burdens to be borne by other users of the same copyright. Economists are not opposed to all subsidies, but where they support them they prefer general subsidies that are provided openly and explicitly and financed by the community as a whole through the tax system, rather than subsidies which are concealed and extracted from some fortuitously chosen group from whom it happens to be convenient to obtain the funds.

---

[1] See P.A. Parisi, "Composition Innovations at the American Society of Civil Engineers," presented at the 1975 IEEE Conference on Scientific Journals, p. 33.

There is a second reason why economists generally do not favor broad exemptions of particular classes of users. As we have noted earlier, a price for the use of some copyrighted material always tends to discourage its use -- the higher the price, the less use will generally be made of the product. Now a zero price to one group of users compensated by a higher price to other users will cause a lopsided change in demand patterns which generally cause a larger overall social loss than a balanced spread of the price increases. There are carefully worked out principles in economic analysis which indicate how the social losses from such distortions can be minimized, and these principles are generally inconsistent with broad exemptions.[1]

G. Compromises Required in Practice

As has already been noted, even if an ideal copyright system can be designed in theory one can be sure that it will require considerable compromise to make it workable in practice. Thus, we noted that while ideally the length of the copyright should vary from product to product, in reality such lattitude could only result in an administrative nightmare.

---

[1] See for example, Y. Braunstein, "An Economic Rationale for Page and Submission Charges by Academic Journals," Center for Applied Economics Discussion Paper, New York University (1976).

At least two other compromises also suggest themselves immediately, though others will undoubtedly arise in practice.

First, the economic principles of optimal charging which were discussed in the preceding section call not only for payments to be borne by all classes of producer, but also for payments to be required for every use. For example, if a charge is imposed upon photocopying of materials protected by copyright, such a charge should, ideally, be required for every such act. Yet (aside from the issue of fair use) with current technology it is literally impossible to monitor every instance of photocopying done in a library or elsewhere. Consequently, user charges for the photocopying of copyrighted material can only be approximated at best. One may be able to impose a flat fee on a photocopying machine or upon a library, or one can base a payment upon the number of recorded uses of a photocopier. However, these all involve three serious compromises: First, there is no way of knowing whether it is the individual who uses the copyrighted material who actually bears the cost. Such a crude payment process must bear as heavily upon the person who copies his own handwritten page as upon the person who copies a journal article to avoid buying that journal. Second, payment will generally not correspond closely to the amount of use of material under copyright. Finally, there is no way of allocating the payments among the producers of the copyrighted materials in proportion to their use. One is driven instead to resort to some sort of pooling arrangement such as the one used to compensate composers of music. Fixed annual payments for performances of recorded music are, for example, made by the television networks to the composers' organization which then divides these proceeds among the composers on the basis of some rule of thumb. This is all a very crude approximation to the theoretical ideal of payment proportioned to use, but enforcement and administrative problems leave no choice except some such compromise arrangement. This type of accomodation may frequently have to characterize payment under copyright in the future.

A second compromise which seems unavoidable is the exemption of very limited usage, for example the photocopying of a single reproduction of a short passage from a journal. The main argument for such an exemption from the viewpoint of the economist is not that such a limited use is "fair", but rather that potential gains from payment for such limited use are likely to be swamped by the administrative and policing costs.

The main issue raised by the necessity of compromises such as have been discussed is whether they should be written into the rules of copyright with an attempt to fix the boundaries of permitted compromise, or whether one prefers to permit time and usage to soften the working rules formulated without exceptions. But this is a choice which lies outside the economist's area of special competence. It is therefore appropriate to turn from this outline of the general issues, to a more careful examination of those directly amenable to economic analysis.

II

IMPLICATIONS OF ECONOMICS OF PRICING FOR STRUCTURING PROTECTION OF COMPUTER SOFTWARE AND COMPUTERIZED DATA BASES

The main purpose of this Chapter is to apply some recent advances in the economic literature on pricing decisions to the question of the scope and nature of property right protection that the society should grant to those who develop and disseminate computer software (CS, hereafter), and to computerized data bases (CDB, hereafter). The major premise that underlies the discussion in this section of the report is that the applicability long-cherished principles of incremental cost pricing to the pricing of CS and CDB is severely hampered for two basic reasons.

(1) The structure of the costs in the STI industry is, in general, such that with prices set at the incremental cost of providing an additional user with the CS package or with the use of a CDB the production costs will not be covered.

(2) Once we grant a property right to the producer we vest him with a certain degree of monopoly power. It is well-known that a monopolist has no incentive to set prices at incremental cost. The question then arises as to the kind of pricing policies that the monopolist should be allowed to engage in.

It should be noted that in this section we abstract from the difficult problem of the duration of the property right. This is the subject of Chapter IV of our report. We begin by stating some fundamentals of pricing and then show how those fundamentals bear on the question of copyright.

A. Economics of Pricing: Some Basic Principles

In the introduction we have already noted some difficulties in applying the long-cherished principles of incremental cost pricing to the pricing of computer software and to computerized data bases. The difficulty that we isolated stemmed from the fact that with price set at marginal cost the total costs of CS and CDS can in general, not be recovered.

This problem is best illustrated with an example. Let us assume that the development, maintenance and updating costs of a particular piece of software, i.e., the fixed costs, are $10,000. The cost of providing the package to an additional customer, i.e., the marginal (or incremental) cost is 10 man hours at $50 per man-hour, or $500. If the firm charges $500, then the initial costs will be uncovered. If the firm charges anywhere above $500, there is a resource misallocation. To see this, note that the cost of serving an additional customer is $500. Consequently, all customers who are willing to pay this amount should be allowed to purchase the commodity. If the price is pushed above $500 because the firm must be able to recoup its total costs, then some prospective customers who are willing to pay the cost of providing them with this commodity will be priced out of the market. Thus

Principle 2.1 If the price of a commodity is set above its incremental cost, some prospective consumers who are willing to pay that incremental cost will be deprived of this commodity.

We should note that *Principle 2.1* applies not only to commodities but also the services that various commodities provide. To illustrate, if the (incremental) cost of accessing one additional entry in the bibliographic data base is five cents, and the price of an access is 5 1/2 cents then all those users who are willing to pay less than 5 1/2 cents will not make the additional search of the bibliographic data base. Unfortunately *Principle 2.1* clashes with the following

Principle 2.2 If the firm is operating at the point at which there are economies of scale in production then it cannot cover its total costs by charging prices equal to the incremental cost.

To a non-economist this principle is not as simple as the previous one. It is imperative therefore that we explain carefully the concept of economies of scale. This is best done with an example. Consider, therefore, a firm that is producing one output, a particular software package. The total cost of providing this CS package to 100 users is $5000. If output (sales) double then total costs of production increase to $9000. Thus doubling of output causes the total costs to increase less than two fold.

The presence of economies of scale can easily be inferred once it is ascertained that in the relevant range of outputs the average cost is decreasing with output. This situation is depicted in Figure 1.

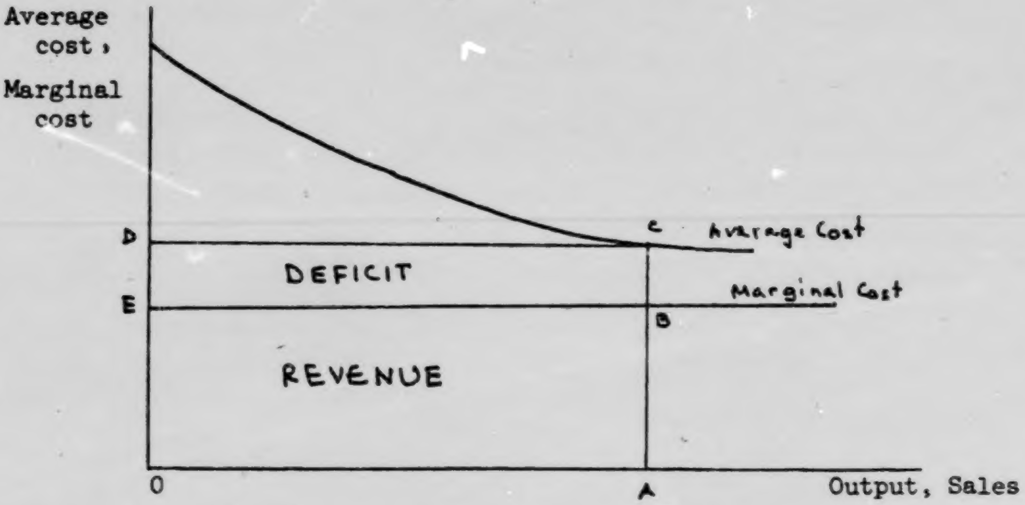


Figure 1

On the vertical axis, we measure in dollars per unit of output the two types of costs: the average cost and the marginal cost. On the horizontal axis we measure output. To fix ideas, the reader may think of output as being the number of sales of a particular CS package. If the firm charges the price equal to OE, its sales are OA. The revenue is given by the area of a rectangle OABE. The total cost is, however, equal to the area of OACD. Hence, the deficit is equal to BCDE.

We have been less than careful in specifying the notion of output. In some cases it may be perhaps relatively easy to specify and measure what the output is. It may be, as in the example above, the number of CS packages sold. Or, in the case of CDB, it may be the number of bibliographic entries extracted from the data base. We shall, however, not attempt to deal with the questions of the appropriate measures of output but rather assume that output is well-defined and measurable.

If we leave the unrealistic world of a single-product firm for the more likely multi-product firms, it becomes more difficult to ascertain those situations in which the marginal cost pricing is not feasible.[1] Here, it suffices to point out that a conceptual difficulty arises because in such firms the notion of the average cost is not well-defined. And as we have seen, the marginal cost pricing is not feasible when, in the relevant range of outputs, the marginal cost is below the average cost. (Luckily, even in multi-product firms, the marginal cost is -- at least in theory -- always identifiable.) Why, then, do we have these difficulties with the average cost in multi-product firms? The answer is that economists, unlike accountants, do not believe that there exists an analytically defensible method of allocating genuine fixed costs among the various products marketed by the firm. Any attempt at allocation is inherently arbitrary. Consequently, depending on the allocation method, some of the products may be priced at the marginal cost while others are not. But a change in accounting practice may lead to the revision of the cost allocations and, therefore in the list of those commodities which can be priced at incremental cost. Fortunately, as we shall see, the economic theory of pricing does not hinge on the notion of the average cost but focuses instead on the marginal costs.

---

[1] This issue is discussed in detail in W.J. Baumol, "Scale Economies, Average Cost and the Profitability of Marginal Cost Pricing," in Essays in Urban Economics and Public Finance in Honor of William S. Vickrey (Lexington, MA: Lexington-Heath, 1975).

Principle 2.3 In a multi-product firm average cost pricing is arbitrary because the relevant average costs cannot be rationally computed.

In order to motivate our discussion later in the chapter on non-marginal cost pricing, we digress for a moment in order to consider the cost structures in various segments of the scientific and technical information industry (STI).

B. Costs in the STI Industries: Empirical Analysis

1. Examples from the scientific journal publishing sector.

Our econometric studies of the cost relationships in the journal publishing sector of the STI industry revealed that both single and multi-product firms generally meet the Baumol tests for the infeasibility of marginal cost pricing. We undertook three separate studies -- one was a cross-section analysis of the publishers of 56 journals; the other two were in-depth studies of large journal publishing organizations.

We present a summary of our findings below.[1]

Our cross-sectional study of journals indicated that, within the sample range, the annual costs of publishing a journal declined as the number of different journal (titles) published by the same publisher increased. This inverse relationship was estimated by a statistical regression analysis of cost and output data of 56 journals from three years. (These data were collected by B Fry.)[2] The regression technique allows us to isolate the effect of the number of journals has on cost from the effects of changes in the circulation and in the number of pages published per year. In particular, a publisher with 36 journals would have average costs of approximately 80% of a single journal publisher after correcting for circulation and number of pages.

The exact equation estimated was

$$\log TC = -.735 + .171 D_s + .588 \log C + .793 \log P - .0621 \log J .$$

In this equation, the journal's total annual cost (TC, in $10,000's) is a linear function of its circulation (C, in 1,000's), the number of pages published per year (P, in 100's), and the total number of journals published by this publisher (J). There is also a "dummy variable" (Ds) which indicates that the costs are higher for journals in the scientific and technical fields. Each coefficient was statistically significant (different from zero) and the R^2 was .8524.

---

[1] Our detailed findings are presented in W.J. Baumol and Y.M. Braunstein, "Empirical Study of Scale Economies and Production Complementarity: The Case of Journal Publication," Journal of Political Economy, October 1977.

[2] Fry, B.& H.J. White, "Economics and Interaction of the Publisher-Library Relationship in the Production and Use of Scholarly and Research Journals", final report, National Science Foundation Grant GN-4398 (November, 1975). Also published by Lexington-Heath.

This cross-section procedure has associated with it several difficulties. The major one is that the per-journal costs from a multi-journal publisher necessarily involve an arbitrary accountant's allocation of fixed costs (generally overhead, administration, etc.). The presence of this problem in our data was confirmed by our finding that statistically there appeared to be no fixed costs; all reported costs could be divided into those varying with changes in circulation and those varying with number of pages.

For example we estimated the following cost function for six journals from one publisher for nine years using standard econometric methods:

$$TC_{it} = -.168 + .184C_{it} + .647P_{it} ,$$

where

TC_{it} = cost of each journal i (i = 1,...,6) for year t (t = 1,...,9) (in \$10,000)

C_{it} = circulation for journal i in year t (in 1000's)

P_{it} = pages published by journal i for year t (in 1000's).

The R^2 was .978. This again shows that costs are linearly related to the circulation and the number of pages. Here the constant term (-.168) was not statistically significant. More formally we could not reject the hypothesis that the true value was zero (at the 95% level).

But we know there are fixed costs in publishing journals. Among the possible explanations one that has a great deal of logic and appeal is that most publishers (or their accountants) allocate these fixed costs among the various titles on the basis of pages, circulation, or some other factor correlated with one or both of them. If true this indicates that we cannot capture all of the economic phenomena in which we are interested unless we find some way to obtain unallocated cost data.

In addition to describing statistically the production process in general, our other interest was in confirming (or denying) the presence of production complementarities in the publication of journals. Here the issue, in non-formal terms, is whether the consolidation of publishing activities or the combining of publishing and other activities results in savings over doing these things individually. The econometric problem was an interesting one: could we develop a specification (a relationship between costs and outputs) that enabled us to measure complementarities and scale economies without, a priori, prejudging whether or not they existed?

One family of specifications that has just these properties is

$$C = a_0 + a_1x_1 + a_2x_2 + a_3(x_1x_2)^{\gamma}$$

where

C = total costs

x_1 and x_2 = outputs, and

$1 \geq \gamma \geq 0$.

Here the a_i and γ are numbers whose magnitudes are to be estimated econometrically.

It can be proved that if a_0, a_1 and a_2 are positive (and $a_3 = 0$), we have scale economies but no complementarities in production. If, in addition, a_3 is negative, these complementarities do exist.[1]

---

[1] See Baumol and Braunstein, op. cit.

We employed this analysis using two alternative sets of data. The data in the first case study came from a publisher with two distinct types of journals -- those reporting original research and those containing translations of articles first published in another language. The cost data were annual costs for each of nine years. The output data were indices that indicated both the magnitude of circulation and the number of pages in the two types of journals. Our estimated cost function was

$$TC = 582.8 + 2.38S_1 + 8.00S_2 - 345.7(S_1S_2)^{0.13}$$

where

TC = Annual costs (in $1,000,000's).

The R^2 was .8012 (but the F was 6.72 because of the low number of observations).

This first analysis confirmed the existence of scale economies and production complementarities. Therefore we can conclude that the consolidation of publication of the original and translation journals resulted in cost savings.

The second case study involved an organization which has affiliated with it a large number of learned societies, each of which publishes a journal as well as providing other services and activities for its members. The data covered fourteen years of operation for twenty-five of the affiliated societies. Using these data we found and measured the economies that arose from the combining of membership and publishing activities.

The estimated function was

$$TC = 153.4 + 1.86M + .680P - 13.17(MP)^{0.1}$$

where

TC = annual society costs (in $10,000's)

P = pages (in 100's)

M = membership (in 1000's)

The R^2 was .794. This confirmed the presence of economies from the combining of activities. This time the range of the observations was much greater than in our previous calculation. The number of pages varied from 20 to 3000 per year and membership ranged from 6000 to 18,000, giving us observations that can be considered to fall very close to the axes and the origin. However, we are less certain of the implications of this cost function because there is necessarily a high correlation between membership and circulation and because we were forced to use number of pages alone to measure the publishing activity of the societies.

2. Scale Economies in Computer Software and Computer Data Bases.

a. Computer Software (CS)

There is an increasing tendency for computer systems managers to utilize standard instead of custom-made computer software. This change has partly been caused by the unbundling of computer software and hardware and has partly resulted from the underlying economics.

> In the first place, standard software generally contains far fewer logical errors; secondly, it is more flexible in that it has been created for generalized criteria, making it easily classifiable and, therefore, readily adaptable to the ever-increasing requirements of the users of data processing;

> it is easily programmable and thus lends itself to shortening substantially the critical phase, rapidly achieving results and, in sufficiently reliable conditions, attaining ultimate objectives.[1]

However, the production of computer software closely parallels the process of publishing in that there are substantial "first-copy" costs that must be incurred irrespective of the number of units produced and sold. If the computer software is a complex operating system, for example, there may also be substantial costs associated with the production and installation of the second and later "copies." However, if these costs decline with each additional unit produced (possibly because of "learning-by-doing") or if some of the pre-production investment is generalizable to all of the copies, there will be economies of scale in the production of computer software.

We were able to obtain some estimates of the development costs of one large on-line retrieval system. This system cost approximately $1,000,000 to develop (including hardware acquisition) over a three year period. This development cost has been divided into five categories. (The approximate percentages of the total follow each category.)

Software Design/Implementation	5-10%
Hardware Acquisition	60-70%
Maintenance	5-10%
Testing & Experimentation	10%
Marketing & Education	10%

---

[1] A. Leggio, "Economic Results of Technical Decisions in Large Data Centers," in A.B. Frielink, ed., Economics of Informatics, (New York: American Elsevier, 1975).

The organization that created it is now marketing this system. The development costs amounting approximately to $250,000 - $300,000 (the sum of the design, testing, and marketing costs) are, at least partially, fixed costs that do not have to be repeated for each "copy". This certainly indicates that there are substantial economies of scale in the production of additional copies or versions of this operating system.

b. Computer Data Bases (CDB)

Here again an analogy with publishing can be shown to be reasonable. The development of a CDB also has large initial costs relative to both the cost of making additional copies and to the cost of periodic updating of the data base. The time profile of the costs of one CDB are shown in Figure 2.

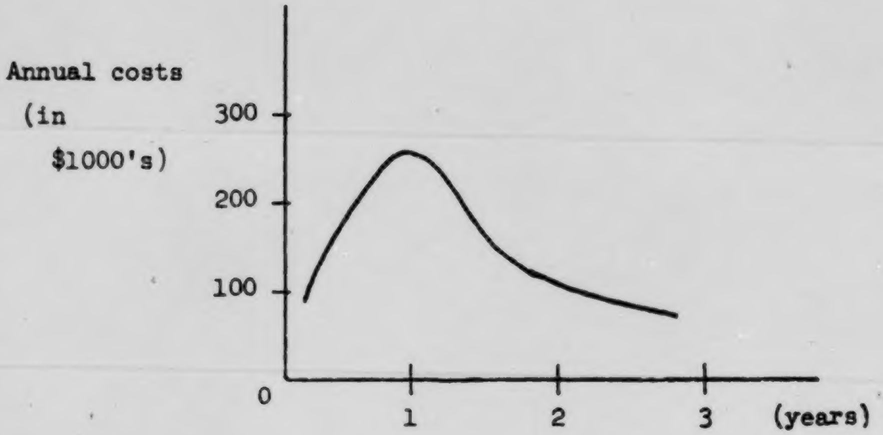


Figure 2

In addition to the scale economies in the production of CDB, there may also be substantial complementarities in production. (Production complementarities occur when it is less expensive to produce two products together rather than separately.) The interrelation between scale economies and production complementarities can be seen in the operations of the abstracting and indexing services. For example, the American Society of Civil Engineers produce printed and machine readable abstracts and indices as part of the same operations. The development costs for this computer-assisted abstracting and indexing system were approximately $25,000. It now costs less than $2,000 per month for the input of 100 items.[1]

The econometric studies reported in this section lead us to the following tentative conclusion: Most STI systems -- scientific journals, CS, CDB -- are characterized by scale economies and in many cases by production complementarities.

C. Setting Prices to Cover the Costs

We have seen already that marginal cost pricing has desirable properties from the standpoint of the society. Thus, if price is equal to marginal cost, all those who are willing to pay the price are per force willing to pay for the resource which is used to provide them with the commodity in question. Ideally, then, we would like to have all the commodities priced in this manner. For a single-product firm, or for a multi-product firm in which all fixed costs can be rationally allocated among various products, i.e., when fixed costs are product specific, the simplistic prescription is to switch from marginal cost pricing to the average cost pricing. By definition, a firm that charges a price equal to or exceeding the average cost can recover all its costs. Setting aside the difficulties associated with the definition of the average cost, we shall demonstrate that average cost pricing has little to recommend it except in a rather unlikely situation which we shall soon describe. Consequently in this section we shall provide some general principles which should guide those responsible for setting prices. The rules to be developed apply to firms that sell an identical product to various categories of users as well as to the firms which sell a variety of products. More importantly the same sets of rules apply

---

[1] See P.A. Parisi, "Composition Innovations at the American Society of Civil Engineers," presented at the 1975 IEEE Conference on Scientific Journals, p. 33.

(a) to a firm which attempts to maximize the welfare of its consumers subject to the proviso that the firm earns an adequate rate of return;

(b) to a firm which strives to maximize its profits.

In order to simplify our exposition of the basic rules, we begin with the very simple case of a firm which sells one commodity. That commodity, however, is bought by two distinct classes of users. For example, a CS package is sold both to profit-making firms as well as to nonprofit research institutes. What distinguishes those two groups is, however, not their goal but the responsiveness of their demands to changes in the price for a CS package. In economics the measure of responsiveness of demand to price changes is known as the elasticity of demand. More formally, the elasticity of demand is given by the percentage change in demand as a result of a (small) percentage change in product price. Mathematically, if Q is the quantity sold and P is a product's price, then its elasticity, e, is given by

$$e = -\left(\frac{\Delta Q}{Q}\right)/\left(\frac{\Delta P}{P}\right)$$

where the symbol Δ stands for change. For example if some price increases from \$3 to \$5 than ΔP is equal to \$2. Similarly, if that price increase causes sales to drop from 10,000 units to 8,000 units then ΔQ is simply 2,000. If e is greater than one it is said that demand is elastic. Similarly, if e is less than one, the demand is said to be inelastic. In Figure 3, we depict two demand relationships which reflect the quantity sold as a function of price.

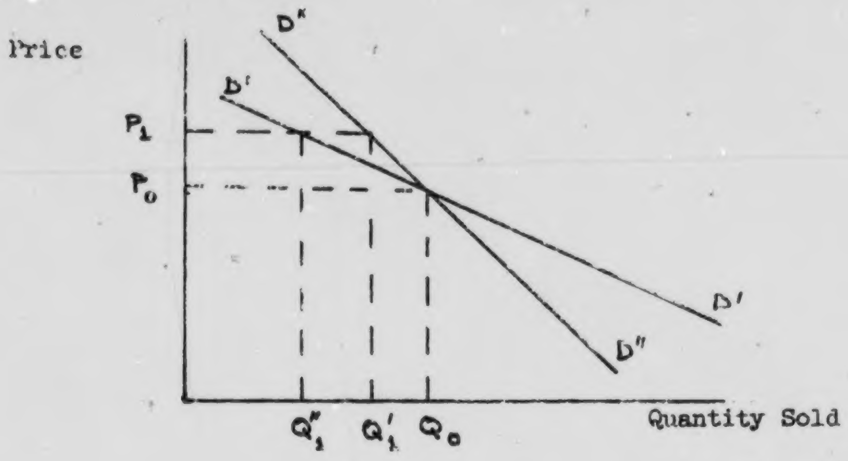


Figure 3

At the price equal to P^0 the same quantity, Q^0, is sold in both markets. However, the demand relationship D"D" is less elastic at the current price than is the demand relationship D'D'. Note that a price increase from P^0 to P^1 reduces the demand in the former market to Q_1 and to Q_1'' in the latter. Clearly Q_1' exceeds Q_1''.

We have argued before that, if feasible, the price should be set at marginal cost. In many markets the forces of competition for consumer's purchases accomplish just that: competition drives down the price to the level of marginal cost. In our example that price is equal to P^0, say. If the firm does not break even at that price, the price must be raised. There will be an undesirable decrease in the quantity sold; this reduction will be most substantial in the markets in which the demand is most elastic.

It turns out that we can give a rather precise measure of the loss to the society caused by a deviation of the price from the commodity's marginal cost. In Figure 4, we plot only one demand relationship.

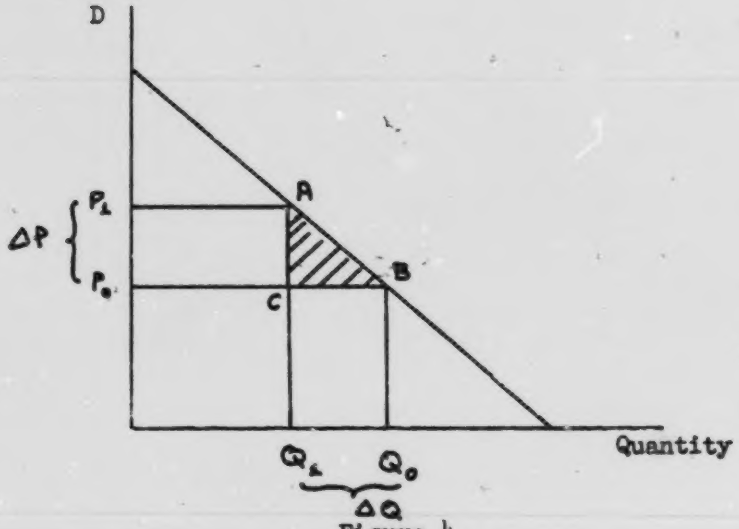


Figure 4

We assume, as before, that the marginal cost is again equal to P_0. It can be shown that, at least under some assumptions, the loss to the society from a price increase can be well approximated by the area of the shaded triangle ABC. In our example this area is obviously given by $(1/2)\cdot\Delta Q\cdot\Delta P$. It measures the loss in consumer's surplus as a result of the price increase.[1]

---

[1] The interested reader is referred to R.D. Willig's "Consumer's Surplus Without Apology," American Economic Review, 66 (September 1976), pp.589-97, for a full review of the concept of consumer's surplus and its uses.

Principle 2.4 Consumer's surplus measures the difference between the maximum amount that a consumer would be willing to pay for a given quantity of any commodity or service rather than to go without it and the actual amount that he pays for that quantity.

To illustrate, with price at P_0 the consumer's surplus is given by the area of the triangle P_0BD. If the price goes up to P_1, consumer's surplus decreases to the amount given by the area of the triangle P_1BD.

It should be already clear that if marginal cost pricing is not feasible then prices should deviate from the marginal cost in such a way that the total loss in consumer's surplus is minimized and at the same time the total cost of the operation is covered. As it is apparent from our Figures 3 and 4 for any given price increase the area of the welfare loss triangle will be the smaller the more inelastic the demand curve. Hence, we have this general rule:

Principle 2.5 If prices cannot be set equal to marginal cost(s), then percentage deviation of price from the relevant marginal cost should vary inversely with the elasticity of demand.

In otherwords, in those markets in which consumers react strongly to price changes, the price should be set close to the incremental cost.[1] However, in those markets in which sales do not respond significantly to price increase, the price should be marked-up substantially above incremental cost. It should be apparent why in the economic literature our *Principle 2.5* is refered to as a Inverse Elasticity Rule of optimal pricing.

---

[1] The validity of this Principle is, however, limited to those instances in which the change in price in one of the firm's markets does not affect the sales in other markets served by the firm. In other words, the rule stated above is valid when there are no cross-elasticities of demand. When these cross-market effects are present some modifications have to be introduced. We shall not discuss those complications here, but rather refer the reader to W.J. Baumol and D.F. Bradford, "Optimal Departures from Marginal Cost Prices," American Economic Review, 60 (June 1970), and J.A. Ordover and R.D. Willig, "On the Optimal Provision of Journals Qua Excludable Public Goods," forthcoming in the American Economic Review. In particular, in the second of the two papers, the authors derive workable pricing rules for the case of interconnected markets.

It is important to realize that a firm which is allowed to price-discriminate, i.e., to charge diverse prices for the same commodity if it strives to maximize its profits it will follow a rule similar to the one just stated: it will change a high price in the inelastic market. The difference between the prices set by the profit-oriented firm and by the firm which has also the interests of the consumers in mind is that the former firm will set prices that will be uniformly higher than those set by the latter.[1] What is important about *Principle 2.5* is that it can serve as a valuable guide to firms since its use will enable them to earn a higher rate of return on their investments.

Charging diverse prices for an identical product is referredto as price discrimination. Some forms of price discrimination are often considered illegal. However, what is not often realized is that

Principle 2.6 Price discrimination may often benefit not only a producer but also consumers. Furthermore, in some situations it is the only way by which a producer can realize enough revenue to induce him to provide a socially desirable product.

We should close this section by pointing out that the general economic theory of pricing as just described, applies to a firm which serves more than two distinct markets as well as to the firm which sells more than one distinct product. In the latter case the marginal costs may, of course, differ from a commodity to a commodity. Still we would find that in a multi-product firm prices should be set in such a way that for each commodity percentage deviation of this commodity's price from the respective incremental cost should vary inversely with the elasticity of demand for this commodity. The percentage deviation can be readily calculated as (price minus incremental cost) ÷ (price). In the "Manual of Pricing and Cost Determination for Organizations Engaged in Dissemination of Knowledge,"[1] the authors provided a survey of various pricing methods that could be effectively utilized by producers of computer programs and data bases to improve their financial position. Without going into the details of the argument here, we may suggest nevertheless that the length of the patent or copyright protection could be shortened if the holder of a patent or of a copyright can reap high returns in the initial period of patent or copyright protection for his CS package or CDB.

---

[1] See W.J. Baumol and J.A. Ordover, "On the Optimality of Public-Goods Pricing with Exclusion Devices," Kyklos, 30 Fasc. 1 (1977), pp. 36-46, [Copy attached].

We shall now turn to ascertaining what are the implications of this general theory of pricing for the proper structure of copyright protection of software packages and data bases. We shall refer to the theory just elucidated as the theory of optimal pricing.

D. The Implications of Optimal Pricing for the Statutory Protection of CS and CDB.

There are important lessons to be learned from the theory of optimal pricing for the proper structuring of copyright laws. We begin by

---

[1] See this Report for more detailed information concerning this argument. W.J. Baumol, Y.M. Braunstein, D.M. Fischer, J.A. Ordover, New York University, April 1977, prepared under the grant SIS74-12785 from the Division of Science Information of the National Science Foundation.

reminding the reader of our basic assumption -- supported by some empirical evidence -- that in the software and data base segments of the STI industry, marginal cost pricing is not feasible because of a significant fixed cost component in the total cost. This fixed cost will have to be recovered from the users/buyers of the commodity.

Let us note that one of the purposes of statutory protection is to secure a return to the developers of CS and CDB sufficient to induce them to provide those products to the market. Consequently, the content and scope of that protection should be such as to secure the necessary return at the minimum dislocation to the efficient allocation of resources. From the standpoint of society that misallocation will be minimized if the prices to all users follow the inverse elasticity rule. This implies, in particular, that the scope of the protection should be quite wide because without such protection it will not be possible to charge the prices that are desirable under the inverse elasticity rule. Specifically it should enable the owner of the legal "rights" to collect payment for the use of his product. To summarize,

Principle 2.7 Unless there are strong reasons to the contrary, the recovery of the fixed cost component of the total costs should be spread over as wide a set of customers and users as feasible. (The contrary reasons include the costs of collection, presence of externalities, and income distribution considerations, and will be discussed in the subsequent sections of this Report.)

The prescription contained in *Principle 2.7* is consistent with the theory of optimal pricing. To see why, let us consider a firm, say a developer of a computer software package, who can sell a CS package outright to the users -- firms or nonprofit research institutions -- who then have free and unlimited access to the package that is stored in the computer's memory. Let us assume, for simplicity, that once the program is stored in the memory it costs nothing to make it available to the user. In other words, the incremental cost of providing the service of the program is zero; the stored computer program is a public good in the sense used in Chapter I. But should the price per use be zero? The answer is no. Note that the fixed cost has to be recovered. This can be accomplished by charging prices, perhaps different, above the marginal cost to the two classes of buyers. (These two prices will be different if the two demand elasticities are different.) The marginal cost that is relevant is the cost of installing and maintaining the package. Now, if a user charge is levied then there will be an inflow of revenue from this source. True, some users will, in general, be discouraged, causing a resource misallocation. However, because of the user charge the two sale prices can be reduced to stimulate the purchases of the package thereby also bringing the price closer to the marginal costs of installation and maintenance. This downward variation in prices will, by definition, improve resource allocation. In otherwords

Principle 2.8 Spreading of revenue collection over a wide spectrum of users reduces the welfare loss which arises from the inability to price the commodity or service at the relevant marginal cost.

The ability to charge user fees is desirable because it spreads the burden of paying for the fixed costs of the package over a wider group of customers. There is, however, little reason to expect that a firm which is granted the privileges of charging user fees will necessarily lower its basic sale prices. Instead, it is likely simply to keep the larger profits that can be obtained from this source. However this result is not unnecessarily undesirable, because it may enable the society to reduce the period of statutory protection. Hence, by allowing the firm to earn greater profits initially the society can reduce the time it waits for the full benefit that will accrue to it from the new computer software package or a data base.

Principle 2.8 The more sophisticated the pricing policy of a firm the higher are the net revenues that it can expect to earn during the period that its product is protected by a property right be it in the form of a trade secret, patent or a form of a copyright.

Incidentally, the annual loss to the community from the monopoly power conferred by copyright can be measured in the same way as the misallocation loss from quasi-optimal pricing. In Figure 5, we plot a demand curve for a service of some product whose marginal cost of production is given by MC^0.

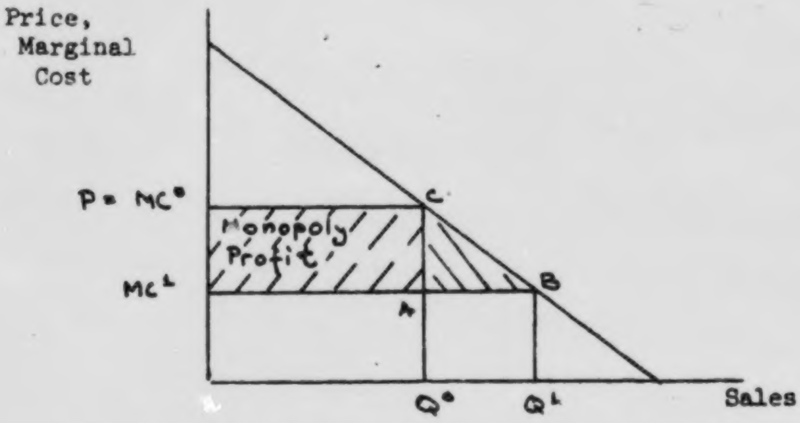


Figure 5

If perfect competition obtains in the production of this commodity, then the price will be set at P^0 and the quantity sold will be Q^0. Let us assume next that by employing some specialized computer program, the firm is able to reduce the marginal cost to MC^1. If it does not have to license its new program to competitors, then for the duration of the copyright it will be able to charge the initial price P^0, keeping all of the savings for itself and passing none of them on to consumers.

The firm's monopoly profits will be given by the area of the rectangle (MC^1) AC (MC^0). Some of these profits will be charged against the costs of developing the cost reducing innovation over which the firm has a copyright. The loss in social welfare per year of duration of the copyright is measured by the area of the triangle ABC. This triangle measures the consumer's surplus that is lost because the firm has monopoly power over the copyrighted cost-saving innovation.

In theory it is rather simple to calculate the social loss due to monopoly power. In practice such calculations are very difficult to carry out. It suffices to point out that estimates of annual welfare losses due to monopolistic pricing practices in the American economy vary from one tenth of one percent of national income to around six percent of national income.[1] There is unfortunately, no easy way to determine the percentage of that loss that can be attributed to monopolistic pricing in the STI sector of the economy. If we assume that total knowledge producing industries in the United States comprise anywhere from 16% to 28% of the Gross National Product, then the welfare loss originating in the STI industries must constitute a rather small percentage of the GNP.[1]

---

[1] See A. Bergson, "On Monopoly Welfare Losses," American Economic Review 63, (December 1973), pp. 853-871; and J. Hirshleifer, Price Theory, Prentice-Hall, 1976, pp. 287-8.

Clearly, the more ingenious the pricing methods employed by the copyright holder, the larger is the consumer's surplus that the firm will be able to extract from the buyers of its products. In deciding, therefore, on the appropriate duration of the copyright, the policy-maker must weigh the benefits of speeding the moment at which the society can capture the full benefits flowing from the cost-saving invention against the costs of extracting substantial monopoly profits during the period throughout which the copyright is valid. As we have seen, the scope for realization of monopoly profits depends on the range of prices that the firm can legally charge.

The urgency of capturing the full benefits of the innovation can be measured by the social rate of discount or time preference. For example, if the social rate of time preference is 10% then a dollar's worth of benefits a year from now has a present value of approximately 90 cents. Consequently, if the discount rate is high, early capture of invention's benefits is particularly beneficial. It is not obvious how such an early release can be accomplished through an interplay of the allowable pricing practices and the duration of protection. Surely, the scope and the duration of protection should be so chosen as to generate a requisite rate of return to the firm. If the duration of protection is short, the firm must be allowed to earn significant monopoly profits early in the period. But the society weighs heavily those initial losses in consumer's surplus relative to the full benefits that will accrue in the future. To reduce those early losses, monopoly profits must be diminished, conceivably by restricting the scope of discriminatory pricing and other special pricing practices. On the other hand if early profits are low the duration of protection must be extended in order to secure the requisite rate of return on investment.

---

[1] The lower figure is from M.U. Porat, The Information Economy, vol. 1 Center for Interdisciplinary Research, Stanford University; the higher from F. Machlup, The Production and Distribution of Knowledge in the United States, Princeton U. P., Princeton, NJ, 1962. Both numbers are reported in Porat, op. cit., Table 4.1, p. 85.

Principle 2.9 The minimization of social welfare loss due to monopoly power can be accomplished by an interplay of the length of protection and allowable pricing practices. For example, if it is necessary that a particular innovation be fully exploited as soon as possible the society may opt for a very low compulsory license fee coupled with the assurance that the developer will have a long period of protection from new products.[1]

It should be noted that our case for user fees does not rest on legal considerations, for example on whether or not the use of a software package constitutes a violation of the copyright.[2] Our preference for user fees stems from the application of optimal pricing analysis. To repeat, the theory calls for positive percentage deviations of prices from marginal costs for all goods and services provided by the firm.

---

[1] For a similar discussion in the area of regulation see V. Goldberg, "Regulation and Administered Contracts," Bell Journal of Economics 7 (Autumn 1976), pp. 426-448. Goldberg discusses the issue in terms of protecting the producer's right to serve.

[2] See "Legal Protection of Computer Software," Computer/Communication Secretariat Report of Working Group No. 12 (September 1976), Canada.

E. Some Extensions of the Theory of Quasi-Optimal Pricing and Their Relevance for the Structure of Copyright Laws

The standard inverse elasticity formula discussed in Section II-C has built into it several simplifying assumptions. The one simplifying assumption that was already discussed earlier stated that there are no cross-market effects of price changes. Unfortunately, the complications that have to be introduced into the formula to take account of the effects of a change in the price of one product upon the demand for another are not easy to describe. Fortunately, the implications for the proper structure of the copyright law in no way depend on those complications.

There are other considerations, however, which may have significant implications for the nature of copyright protection. The first involves the problems of income distribution. We observed earlier that the welfare loss from monopoly pricing can be fairly well approximated by the triangle ABC of Figure 5 which measures the dimunition in consumer's surplus resulting from "incorrect" prices. Now, the decision-maker may well consider the loss in consumer's surplus in one market to be more serious than an equal loss in some other market in which the commodity -- a CS package or a CDB -- is sold. This differentiated evaluation should lead the policy-maker to reduce the price in the market in which losses of consumer's surplus are considered highly detrimental and to raise the price in the remaining market or markets. For example a new computer routine can be used to improve the manufacture of various commodities bought mostly by low-income people as well as to lower the costs of some luxury commodities. If the demand for luxury commodities is more elastic than that for the necessities the profit maximizing firm which follows the inverse elasticity rule (see *Principle 2.5*) will set a high price in the latter market and a lower price in the former market. However, for reasons of equity the society in general will value more the loss of a dollar in benefits to the low income groups than a loss of a dollar's worth of benefits to the high income groups. Consequently, from society's standpoint welfare loss figures have to be modified by correction factors designed to reflect the policy-maker's concern about income distribution.[1]

The profit-maximizing firm, however, is not concerned with the niceties of differential evaluation of welfare losses. It focuses on elasticities of demand and sets prices in accord with what the traffic will bear. But the policy-maker can introduce incentives which will cause the profit-oriented firm to respond, at least partly, to those differentiated evaluations. The simplest route, perhaps, is to reduce the period of copyright protection for commodities sold in the preferred markets. Such a move would have to be accompanied it would seem, by a compensated increase in protection elsewhere in order to guarantee the rate of return judged necessary to stimulate the developed process. A reduction in the period of protection in the socially favored segment of the market, holding the flow of new products constant, would increase the pool of substitutes which would be available at low (or zero) prices. This increased competition would tend to increase demand elasticity for the CS package or CDB that has just come on the market. We have seen already (cf. *Principle 2.5*) that an increase in the elasticity in demand reduces the profit maximizing percentage mark-up of a price over the relevant incremental cost. Consequently such policy may have the desired effect of lowering the price in that sector of the market in which it is important from a social standpoint that the loss in consumer's surplus be held down to the minimum.

---

[1] See R.D. Willig and E.E. Bailey, "Ramsey-Optimal Pricing of Long Distance Telephone Services," Bell Laboratories Economic Discussion Paper #77, January 1977.

Unfortunately the policy prescription just stated may suffer from two perhaps fatal difficulties:

(1) It is not obvious that the flow of new products will not be adversly affected by the reduction in the length of protection.

(2) The compensatory extension of the length of protection in the socially less-favored segments of the market for CS and CDB may prove to be of little economic value.

Let us take up each of those points starting with the second one. The value of an extended life of the property right depends on the additional profits that can be earned because of this extension. If the economic value of the property right expires before the end of the period of legal protection then the extension is valueless. There is some evidence that the economic life of a new product is much shorter than the period of legal protection, particularly in the rapidly developing segments of the STI industry such as computer software. The average economic life of computer software has been estimated to be around three years, which is substantially less than the proposed period of legal protection.[1] However, there are two additional considerations: First the variance may be high; some CS packages (e.g., BiMed, SPSS, etc.) may have relatively long economic lives. In testimony to CONTU,

---

[1] Canada Department of Consumer and Corporate Affairs, Working Paper on Patent Law Revision, Ottawa, Canada, June 1976.

Lockheed estimated the economic life of the DIALOG operating system to be on the order of ten years. Second, the documentation can have an independent copyright thus enabling the creators to obtain payments from the users. It is likely that these two considerations go hand-in-hand; i.e., CS with longer economic lives may be the CS for which the documentation can be independently sold. Nonetheless, it may be that the "long-lived" CS's are just those we seek to protect because of their usefulness to society. Therefore a short average economic life is not, per se, an argument for a shortened period of protection.

It is also arguable that a manipulation of protection length in the various sectors of the market for CS and CDB may affect the flow of new products. In addition, there may be substitution in the products' design so that the products would be more suitable for use in those markets where competition is less strenuous. This long-run effect -- the alteration in the flow of new products -- can reverse the short-run, transitory benefits which may accrue to some users because in the short-run the flow and composition of the new products will not be significantly affected.

This, incidentally, provides a guarded argument against various exemptions that have been contemplated in the copyright law. Those exemptions simply exclude various groups from contributing their share of the total cost of producing and disseminating the product. The proponents of such exemptions must implicitly assume that there are unexplored profit margins in the other segments of the market for STI in which the producers can make-up their lost revenue, and that there is a clear case for this sort of cross-subsidy by the one type of consumer by another. Thus we have this loose economic rule:

Principle 2.10 Exempting some users from payment or arbitrarily restricting the charges that can be assessed against them leads to increased costs and charges to the less-favored users.

Similar remarks apply to the proposals for compulsory licensing. Under compulsory licensing, the prices charged in all markets may well turn out to be uniform. Such pricing is not necessarily desirable either from the standpoint of profits or social welfare measured by the size of consumer's surplus generated in the industry. This follows from our previous analysis which revealed that some form of price discrimination is preferable in general to uniform prices. The effect of compulsory licensing serves to reduce monopoly profits but the undesirable long-run effect of such licensing may be a reduction in the flow of new products to the market place.

Principle 2.11 Compulsory licensing and other artificial restrictions on prices reduce the welfare loss from those products that reach the market. However, the total effect of such restraints may be undesirable if it impairs the flow of innovations.

We close this chapter of our report with a conclusion, albeit tentative one, that the protection afforded to the producers and disseminators of CS and CDB should not include unreasonable restrictions on the admissible pricing schemes. If the pricing practices employed by the producers are found undesirable they should be attacked directly, using the existing antitrust laws. In addition, if it is found that legal prices charged by the producers are undesirable for reasons of income distribution or externalities that are generated by some buyers of CS and CDB, then those problems should be also attacked directly, through subsidies, for example, of some classes of users or producers.[1]

F. Transactions Costs in Computer Information Systems

We have noted previously that the likelihood that transactions costs will be high can lead either to a decision against user fees, in general, or to the exemption of certain classes of users from payment because the volume of usage by them is so small that it is not practical to levy a charge. The economic issue is the comparative efficiency of user charges versus free provision and the relative magnitude the transactions costs. For this comparison to be valid it is necessary that all the accounting, billing, collection, and enforcement costs be included as transaction costs.

The level of these costs will depend on the coverage required and on the form of the license. For example, if infrequent use in personally owned mini-computers is exempted, the transactions costs for the collection of royalties for the use of some computer programs may be substantially reduced. Similarly, blanket licenses generally result in lower accounting costs for one or both parties because detailed information on the level of usage is not required.

Operators of two different types of computer systems provided us with estimates of the costs of monitoring of use, accounting, billing, etc. One was a large "service bureau" organization with a large number of customers. They estimated that their transactions costs were generally 15 to 18 percent of total costs. The other organization operated only an on-line information retrieval system. Their costs of monitoring and recording usage were approximately 10 to 15 percent of their total system costs. However, it is important to note that in neither case do these costs include the expense of determining how to disburse any royalties or fees due to copyright owners.

---

[1] See J.A. Ordover and R.D. Willig, "On the Role of Information in the Design of Social Policy Towards Externalities," Center for Applied Economics Discussion Paper #76-03, New York University, for a guide to various methods of coping with external effects.

The obvious archetypes of a clearinghouse that collects royalties and then makes payments to the copyright owners are the performing rights societies. ASCAP usually has been able to hold the level of its operating costs below 20 percent of its revenues. The disbursement costs are a function of the payment scheme used. The more detailed the information required (by monitoring or some reporting plan) the more expensive it will be for the parties involved. In general, usage of a variety of goods or services such as copyrighted works does not follow a uniform distribution. In fact, usage is usually highly skewed. To use an example from the photocopy field, the British Lending Library reported that of the 14,967 serial titles in their collection, the most requested 210 titles accounted for 20 percent of the demand for photocopies while the 6000 least requested titles amounted to only 10 percent of the demand.[1] This sort of skewness can lead to a substantial reduction in monitoring and other transactions costs if there is some minimum level of usage for which no royalty payments are distributed. On the other hand, the absense of such a "threshold" may require quite accurate and complete monitoring to insure the detection of the smallest levels of usage.

---

[1] M.B. Line and D.N. Wood, "The Effect of a Large Scale Photocopying Service on Journal Sales," Journal of Documentation 31:241 (1975).

III

THE CONSEQUENCES OF TRADE SECRECY AND COPYRIGHT PROTECTION FOR THE QUALITY, VARIETY AND QUANTITY OF NEW PRODUCTS

In the last chapter we discussed some implications of economic theory of pricing for the structure of property rights that are to be granted to producers of software and data bases. In particular we found that there exists a trade-off between the length of protection afforded the producers and the breadth of this protection in terms of allowable pricing practices. We also argued that there are strong reasons for not interfering with the price mechanism through the copyright laws but rather for leaving the stamping out of the undesirable pricing practices to the authorities responsible for the enforcement of the antitrust statutes. Specifically we tentatively concluded that if copyright protection is in the end extended to the producers of CS and CDB they should be entitled to sophisticated user charges and that no classes of users should be exempt from payment. We also noted that compulsory licensing may have undesirable consequences for the long-run flow of new software and data bases.

The discussion in the last chapter implicitly assumed that some form of statutory protection will be in place. In this chapter we examine the differential impacts of trade secrecy and copyright protection on the quality, variety and the quantity of new commodities that will be available in the marketplace for the computerized scientific and technical information.

Specifically we shall find that trade secret protection biases the quality of new output in the direction of unnecessary complexity. It also channels the innovative efforts into those forms of software which are not particularly well-suited to general use. In other words trade secret protection is particularly detrimental to private provision of non-specialized, general interest types of software. Copyright protection diminishes that type of distortion. Certainly under compulsory licensing the producers would strive to provide the product with as wide a applicability as possible.[1]

There are no general conclusions to be drawn as to the quantity of the new products that would be forthcoming under various regimes of property rights as compared to the socially optimal quantity. Copyright protection may be expected to intensify competition in the software and data base industry, vis á vis a situation that would be obtained under trade secret protection. It has been shown by various authors that competitive structure of the industry is not necessarily most conducive to innovative effort. In fact the purpose of copyright protection is to transfer a competitive situation into a (temporary) noncompetitive one.[2]

Protection based on trade secrets may lead to larger concentration in the software and data base industry which may be more conducive to the optimal flow of new products. It is well known, however, that the presence of trade secrets does often lead to significant duplication of effort because of informational imperfections.

---

[1] It is often argued that if compulsory licensing is instituted the unusual tastes may go unsatisfied. For example some say that if General Mills were to have to license the production of Cheerios it would optimally adjust the receipe to conform to more standard tastes.

[2] For a recent mathematical study of the problem of the interrelationship between market structure and the flow of new products see G. Lowry, "Market Structure and Innovation," The Center for Mathematical Studies in Economics and Management Science, Discussion Paper #256, Northwestern University.

It is obvious, nevertheless, that the flow of products will be directly related to the expected monopoly profits that the innovator hopes to realize from the product. We have noted in the previous sections that the welfare loss during the duration of the monopoly is the price that the society must pay in using copyright as its inducement for investment in the development of new products. The longer this period the more such investment it should stimulate, but the greater the interim loss in social welfare. Weighing these two factors against one another one must use them to determine how long this monopoly should last. In part, the question may be somewhat academic in that the economic value of the copyright may evaporate much sooner than the copyright itself. For example, the economic life of a software package has averaged some three years -- much less than the prospective length of the copyright protection. (See our discusison of the relevance of "average" economic life in Chapter II-B above.)

The question of the optimal length of protection can also be rendered academic if the society opts for trade secret protection as opposed to the statutory protection. Under a trade secret approach the duration of effective protection is in fact a random variable whose realization depends on the success of the possessor of the secret in preventing it from falling into the public domain. We do not, of course, have any data on how successful the possessors of trade secrets have been in securing their monopoly power. However, it is an interesting fact that the holders of patents have had a difficult time in defending their property rights.

As Professor David Bender notes, "statistics indicate that in about 70 percent of infringement suits litigated to conclusion, the defendant emerges the victor."[1] Although Bender realizes that patent protection is inherently uncertain he seems not to note that the same applies to trade secrets. Unfortunately, we do not have comparative data on the riskiness of the two forms of protection. The substantial investments that firms make to protect their secrets do indicate, however, that the probability of losing trade secret protection is rather substantial.[2]

The expected probabilities of losing protection under the two distinct,regimes -- trade secrets vs. statutory protection through patents or copyrights -- figure quite significantly in an evaluation of the virtues of the two systems. The main criticism of trade secrecy is that, because it depends on the prevention of disclosure, it inevitably leads to a wasteful duplication of effort. There is no need to recount here the horror stories that are being told about the duplication [illegible] the programming work in many corporations. The costs of duplication and the attendant waste must be weighed against the benefits of the probability of quick disclosure at a zero price to the general public. Thus, before the idea of increased reliance on protection based on trade secrets is discarded those costs and benefits must be ascertained. The general presumption is, nevertheless, that the trade secrecy method of creating incentives for innovation is inefficient.

In the next two sections of this chapter we examine the structure of incentives for innovation under various regimes of property rights.

---

[1] D. Bender, "Trade Secret Protection of Software," The George Washington Law Review, 38 (July 1970), pp. 909-57, 915.

[2] See Bender, ibid., pp. 912-13.

We postpone until Chapter IV the discussion of the optimal duration of protection.

A. Qualitative Effects of Trade Secrecy and Copyright Protection

The producers of products have not only the choice of how much to invest in the research, development, marketing, etc., of new products but also a choice of characteristics of the products that will be developed. The effects on the choice of the product or products are classified here as qualitative effects of protection. In other words, the nature of the protection that is allowed for intellectual property will affect the quality and the variety of the commodities produced in the economy.

Trade secrecy protection favors developments of those products which cannot easily be duplicated or engineered backward. Consequently, under such a regime, one is likely to find that producers will build into their products various special features that may complicate appropriation of the secret by unauthorized users. For example, there may be an unnecessary tendency for complication of the structure of the product, so as to obscure its essential properties and increase the costs of unauthorized use. In computer programs there may be a need for introducing special safety devices.

A typical list of features of computer programs which are designed so as not to be usable by outsiders is the following:

(1) Frequent use is made of special features of the machine on which the program is currently running (such as machine language instructions, features which depend on the word length, on the representation of characters by bit-strings, etc.)

(2) All comments, which may make the logical flow of the program clear to a reader are omitted.

(3) The loop structure of the program is unnecessarily complicated, with transfers of control back and forth within the code.

(4) Mnemonic variable names, which could reveal the meaning of a variable are avoided.

(5) Unnecessary portions of code are added to confuse a potential user.

It should be clear that all these features reduce the social value of a program or software package, and do not only make it more difficult, or nearly impossible, for an outsider to use the program or appropriate its logic, but also make it harder for the original programmer to maintain and modify the program. But many programs are, in fact, written in the manner just described to protect them against non-paying users.

Principle 3.1 Trade secrecy protection leads to the waste of social resources through overly complex product design. Another distortion that may occur is that programs and data bases will be tailored to the needs of specialized groups of users. By making the program more specialized the developer minimizes its usefulness to others and diminishes the probabilities of disclosure. To borrow language used in the economic theory of investments in human skills, trade secret protection makes specific investment more valuable than general investment.[1] General investment is to be understood as the investment in the development of programs and data bases which can be employed by a wide variety of users.

---

[1] See G.S. Becker, Human Capital (NBER and Columbia University Press, New York, 1964).

However, the more widespread the use the more difficult it is to prevent the leakage of the technical desscription of the products into the public domain which, in turn, undermines the protection afforded under the law of trade secrets. This qualitative variety bias of trade secrets has not been discussed in the literature. It seems, however, that it may be a cause of even worse misallocation of resources than the effort duplication commonly viewed as the most serious consequence of trade secrecy protection. To minimize this bias, direct support on contractual basis may have to be provided to those institutions which may be expected to develop the general purpose programs or data bases.[1]

Principle 3.2 Trade secrecy protection may lead to overprovision of specialized CS and CDB and to underprovision of general-use CS and CDB. Another form of the quality-variety distortion occurs when, for maintainance of secrecy, the maintenance, revisions, and so on, of programs and data bases is undertaken by the developer of the CS package or CDB even though efficiency would dictate that this task be carried on to some extent by the user himself. For example, we find that computations in the structural analysis of buildings or airplanes are carried out by the owner of the program rather than by the firm which needs the calculations. It is not clear whether such an arrangement is optimal, for it necessitates a significant flow of information between the user and the owner. This type of use of a program, in which the user supplies the data and receives the final results, but is not allowed to put his hands on the program itself, is particularly inconvenient when the program in question ought to be part of a larger model. For example, it is impossible to apply an optimization routine, which the user may have available on his own computer installation, to a structural analysis program which he cannot transfer to his computer, but for which he has to supply the input data for each run in prespecified form to the firm who owns the program.

---

[1] The economic analysis of this type of governmental support is provided in J.A. Ordover and R.D. Willig, "On the Role of Information in the Design of Social Policy Towards Externalities," Center for Applied Economics Discussion Paper #76-03, New York University.

Lastly, we must allude to the fact that various forms of protection may cause greater or smaller reliance on bundling and tie-in sales. Bundling and tie-ins are widely used as a method of indirect collection of payment for a commodity for which exclusion of nonpayers may prove to be difficult. Thus, in the brokerage industry, research was routinely provided "for free" as part of services generally performed by stockbrokers.

With computer software, bundling may take the form in which a producer of a software package makes its use available "for free", provided the user runs the program on the producer's computer and purchases computer time simultaneously, at an inflated rate. One would expect that unbundling -- perhaps partial -- would be a consequence of greater reliance on copyright or patent protection of CS packages and CDBs. The possibility of unbundling arises in the regime of copyright protection because the sellers are not forced to collect payment through roundabout routes. Previously bundling and roundaboutness were introduced to reduce the fear that the product will fall into the public domain and lose any protection from the nonpayers.

Principle 3.3 Copyright protection may reduce the efforts of the sellers of software and data bases to bundle their products with other products or services for which the costs of collecting the sale price or user fee are low. However, there would probably be a continued reliance on tie-ins for the purposes of extracting more profits from buyers of these products. We have discussed the advantages and disadvangages of tie-ins elsewhere and there is no reason to review our findings at this point.[1] Here we simply want to point out that bundling and tie-ins can be used as devices to increase the transaction costs of unauthorized use. Whatever the reason for the use of bundling, there will be a gain in profits and, in the short-run at least, losses in that component of social welfare that is measured by consumers' surplus.[2]

Statutory protection through copyright or patents has as its cornerstone the disclosure of the fundamental properties of the product. With (more or less) full disclosure there will be no special need to design new programs to have features which increase the transaction costs of unauthorized appropriation. This leads to the hypothesis that statutory protection may spur the development of general use programs. Under statutory protection, disclosure of the characteristics of the product to the general public need not have significant effects on the profitability of any given product. As a matter of fact, the demand for the product -- a software package, for example -- is generally stimulated by knowledge of the product's availability. But such knowledge can be detrimental to profits. Improvement in the information about products increases competition for prospective buyers who now can make better choices. Competition, of course, drives down the rate of return on new commodities. Consequently, our hypothesis will be valid if the reduction in profits due to better consumer information is small relative to an increase in expected profits due to a reduction in the probability that the CS package will be appropriated without payment.

---

[1] See W.J. Baumol, Y.M. Braunstein, D.M. Fischer, and J.A. Ordover, "Manual of Pricing and Cost Determination for Organizations Engaged in Dissemination of Knowledge," New York University, April 1977.

[2] Bundling adds flexibility to the pricing schemes to be used, and in this way offers a potential to increase welfare in a way similar to price discrimination, which was discussed earlier. If bundling is used by an unregulated monopoly, it will always increase profits. It may increase or decrease consumers' surplus. An example where price bundling increases consumers' surplus (and therefore social welfare, since it also increases the producers' surplus) is a case in which a software package will be developed only if the producer can recover costs through price bundling, but not otherwise. The users' willingness to pay for the package must exceed their actual payments, by definition, since they are not forced to buy. If the package could not be produced at all, because of a prohibition of price bundling, then the consumers' surplus would be zero in this particular example. But there are also many cases where the use of price bundling reduces social welfare, as shown for example by W.J. Adams and J.L Yellen, "Commodity Bundling and the Burden of Monopoly," Quarterly Journal of Economics, XC (August 1976), 475-98.

It is implausible that statutory protection will have an adverse effect on the flow of specialized programs. Disclosure should not be harmful to those products. They are intended, to begin with, for a small group of users and in the limit for only one user. In those situations, the value of the product to others may be quite low and not worth the price of "cracking" its structure.

Principle 3.4 We tentatively conclude that a shift from trade secrecy protection to statutory protection should have stimulating effects on the variety of the products developed and marketed by the software industry. We expect that under the latter form of protection, private firms will have stronger incentives to produce general use software packages without substantial funding commitments from the government.

Funds for the development of general purpose software packages have occasionally been provided by the government and by private foundations.

For example, the Time Series Processor (TSP), an econometric package, was developed at the Harvard Institute of Economic Research with the support of the Energy Policy Project of the Ford Foundation, the Office of Emergency Preparedness, and additional funds provided by the participating academic institutions. NASTRAN, a program for use in structural analysis, was developed by NASA and then made available without charge to any potential user. Many general purpose algorithms have also been written and published by academic researchers who performed sponsored or non-sponsored research. A representative collection of these is contained in "Collected Algorithms of the Communications of the Association for Computing Machinery," which is updated regularly and is available to potential users for a small subscription fee that merely covers printing costs. One hopes that with a clearer form of protection for software, private industry may begin to develop the general use programs the development of which, until now, relied heavily on government subvention.

B. Quantitative Effects of Trade Secrecy and Copyright Protection

In the previous section we noted that the mode of protection will affect the variety of products in the marketplace. It is equally clear that the choice of protection systems should affect the quantity of the products. Thus, it is generally known that a monopolist will choose not only inappropriate product characteristics but given the chosen product then will underprovide it to the public. In the last section we have seen that the reliance on trade secrecy may cause the (prospective) developer to shy away from products that are intended for wide circulation because the costs of enforcing the trade secret are high for those types of products.

Abstracting from a product's characteristics, we may inquire about the supply of products under various forms of protection. The general result which emerges from theoretical discussions is that irrespective of the kind of protection, competition will tend to overprovide inventions and innovations with high private value, and simultaneously underprovide those with high social value. These are two opposing forces, one of which tends to increase the supply of inventions over the socially optimal level, and another tendency which decreases the amount. Which effect will outweigh the other depends on the specific situation. There will be an increased supply of inventions and innovations which bring a private advantage to the supplier, possibly at the expense of the rest of society. On the other hand, inventions which are useful to society as a whole, but difficult to sell and therefore of relatively small value to the producer, will be supplied in a less than socially optimal amount. For example, programs which imitate or duplicate existing programs and therefore add very little to social welfare, but which enable the producing firm to capture a portion of the corresponding market and enter a lucrative business, will tend to be overproduced. But programs of a highly innovative nature, which could be beneficial for a wide variety of users for a long time period, but which it may be more difficult to market, will tend to be underproduced. The net effect can be an increase or decrease in innovative activity from the socially optimal level, but it is clear that with competition the wrong kind of innovation will be forthcoming.[1] It is also clear that with competition a larger amount of resources will be devoted to the production of any given amount of innovation than is socially necessary.

---

[1] For a discussion of these ideas, see J. Hirshleifer, "The Private and Social Value of Information and the Reward to Inventive Activity," American Economic Review 61 (September 1971), pp. 561-574.

Under trade secret protection, the dangers of overprovision are magnified because there is lack of information as to the existing products. Overprovision in the regime will take the form of too many identical products being developed by independent producers. As we have seen, however, the variety may also be affected adversely.

Although trade secrecy may lead to an overprovision of new computer software and data bases, because of a lack of communication, it also may lead to excessive concentration in the industry, which in turn brings with it the monopolistic tendency to restrict output in order to maintain higher prices.

The proposition that there will be overprovision under competitive market structure depends on the assumption that the number of firms which engage in the development of new computer software and data bases is fixed. However, as we noted above, the use of trade secrets can lead to excessive concentration in the industry. The argument can be summarized as follows: Lack of information about existing CS packages places a premium on the control of a large portfolio of those packages whenever there are some complementarities between various items of software. In the literature, those gains are referred to as economies of scope. Economies of scope exist whenever the total costs of producing two or more items of software in separation exceed the total costs of producing them in one firm.[1]

---

[1] See John C. Panzar and R.D. Willig, "Economies of Scale and Economies of Scope in Multi-Output Production," Quarterly Journal of Economics, (forthcoming), for a complete discussion of the concept.

The complementarities between various items of software may be attributable, for example, the interchangeability between subsets of computer programs. If a firm cannot obtain the information about other programs which can be useful in the development of the new one then it will have to waste resources on reinventing the existing complementary software. A firm with a large portfolio of programs is at a compartitive advantage over a small firm with the proprietary programs. If copyright protection is substituted for trade secrecy then the flow of information is facilitated and firms can fully utilize the existing stock of knowledge. Hence

Principle 3.5 Barriers to entry, and hence competition, will be diminished by a move from trade secrecy to copyright protection.

Disclosure also facilitates competition because the downstream firms -- the purchasers of software -- have better knowledge of the available products. An improvement in consumer information should drive down the price. More important, it may also ease entry into the upstream -- software -- industry. Then the argument is that with better information the downstream firms can more easily search out alternative sources of supply. Hence, the well-established firms need not have a special advantage over the newcomers because the newcomers can advertise widely and specifically their products without the fear of losing trade secrecy protection. This discussion brings out a point which curiously seems to have been missed in the current literature, namely that disclosure not only reduces wasteful duplication of the research effort, but also reduces search costs for the ultimate buyers. This reduction in search costs facilitates improved matching between buyers and sellers and improves social welfare.

Principle 3.6 Copyright protection reduces the transaction costs of matching buyers and sellers relative to those costs that would be incurred under policy of trade secrecy.

There is no agreement in the literature on innovations on the sign of the correlation between the market structure and the flow of innovations. The Schumpeterian view is that monopoly leads to improved flow of research over time. It adds credance, therefore, to those policies which lead to market concentration in the upstream industry. That is, monopolization yields an improvement in intertemporal allocation of resources. On the other hand, monopoly has detrimental consequences for static or temporal efficiency: As we showed earlier, it leads to a loss in consumer's surplus. (There may be also a loss in producer's surplus if the marginal costs are not constant over some range of output.)[1] Consequently, if our argument that trade secrets lead to market concentration is correct, then the gains in intertemporal welfare which arise from increased concentration and the reduction in the duplication of research effort should be measured against the known static (temporal) costs of monopoly.

We found in this chapter that there are strong arguments in favor of copyright protection in terms of the quality and variety of output of software and data bases. Trade secrecy protection leads to overprovision of specialized types of software and CDB's relative to the socially optimal level and to the concommitant underprovision of general, widely applicable software and data bases. Those tendencies are attenuated by disclosure policy which is an essential component of copyright protection.

---

[1] See M.I. Kamien and N.L. Schwartz, "Market Structure and Innovation: A Survey," Journal of Economic Literature, 13 (March 1975), pp. 1-27.

We may also conclude that trade secret protection leads to the concentration in the STI industry because it creates artificial barriers to entry into the industry and impinges on the efficient matching of buyers and sellers.

To the extent that competition is baneful to the provision of new products as suggested by the Schumpeterian hypothesis then the monopolistic tendencies encouraged by trade secrets should be welcomed. However, the benefits of improved flow of innovations have to be measured against the quality distortions and transaction costs caused by trade secrets protection.

In the next chapter we shall study the appropriate length of protection on the assumption that such protection does not exceed the expected economic life of the property.

IV

THE OPTIMAL DURATION OF COPYRIGHT

A. Factors Which Influence the Optimal Duration of the Monopoly

It has been mentioned before that in order to induce firms to produce computer software,[1] some form of protection has to be granted to them. For if every computer program or data base were to become public property as soon as it was developed, the firm which invested resources in its development would hardly be able to recover its costs, and would consequently have very little inventive to undertake such research activities. A temporary monopoly on the rights to use or sell a piece of computer software will give a financial reward to the successful innovator, which encourages him to undertake such activities. The longer the period for which such a monopoly is granted to an innovator, the greater will be his incentive to devote resources to the development of computer software, and the more abundant will be the stream of new programs and data bases being developed.

But a long period of protection also has an adverse effect on public welfare. The longer a monopoly is granted to an innovator, the longer the period will last during which there is a misallocation of resources. The sooner restrictions on the use of a computer program or data base are lifted, the more useful will that particular item be to society, because, as discussed earlier, the cost of additional uses is likely to be negligible.

---

[1] The same analysis applies to innovations in general. For a more thorough discussion of these issues, see "Optimal Lags in a Schumpeterian Innovation Process," by W.J. Baumol and D.M. Fischer, mimeo, New York University, 1976.

The problem is, then, to find an optimal compromise between the two extremes of too short or too long a lag period during which protection in the form of a copyright is granted to the producer.

In the next section we construct a model that shows how a firm reacts to various periods of protection. In that model we will assume that there are diminishing returns to scale in the production of computer software for two reasons. First, we assume that there is an unlimited number of potential projects which can be undertaken by a computer software firm. Some of them promise very high benefit to cost ratios, while this ratio will be lower and lower for the less attractive projects. The firm will undertake all those projects for which its expected benefits exceed its expected costs. Here the benefits to the firm consist of both the gains it derives from its own use of the computer software it has developed, and of the revenues from sale of the program or from the licensing of its use. Costs will include a normal rate of return on invested capital. They will also include some premium for risk taking, because of the risky nature of any innovative activity, such as the development of a new computer program, where it is difficult to estimate in advance the difficulties and obstacles to be encountered and consequently the number of manhours to be expended. In many cases, it may even prove impossible to complete successfully a project that has been started, so that the investment will turn out to be a complete failure. Riskiness of this sort is typical of any kind of innovative activity, for example the development of new drugs or new products in general. Incidentally, in dealing with this problem larger firms, which can pool risks, have an advantage over smaller firms. Small for-profit firms will tend to engage in the development of computer software that follows more or less established lines and has direct applications, whereas the more risky basic research in new computer languages and algorithms will be left to research units within large firms, or to publicly supported academic institutions.

A second element which causes returns to diminish with increased activity is the considerable variation in the skills of computer programmers. One brilliant programmer may well be able to accomplish more than twenty mediocre ones.[1] As a firm expands its operations, it will first tend to hire the most skilled personnel, and gradually have to resort to less and less talented employees. These two factors, the declining benefit to cost ratio of software projects considered for implementation, and the declining abilities of incremental personnel, tend clearly to lead to diminishing returns to scale in the production of computer software, at least at the level of the industry as a whole. Although these will be some cost advantages in concentrating the development of computer software in a few large firms, which can benefit from internally shared information and reduced risk, every firm will operate at a level where it encounters increasing costs per unit of output with expanding activity.

Obviously, the longer the period of property right protection that is granted to a firm, the higher will be its revenue from any particular project it may undertake, and the more profitable, on the average, that innovative activity will therefore turn out to be. On the other hand, a langer period of protection will reduce the availability, and with this the usefulness to society, of any given amount of computer software produced. The model in the next section will permit us to calculate the optimal length of the period of protection under given assumptions about the shape of the cost curve and the structure of demand for computer software.

---

[1] Although programmers with greater skills usually command higher salaries, the differences in pay will usually hardly be commensurate with the much greater differences in productivity. But even if programmer salaries were to capture all of the benefits of superior productivity it would still be true tautologically that the best programmers can yield the largest product. All that changes in this case is the identity of the recipient of the benefit.

On the demand side, we assume that the amount users are willing to pay for any given computer software package declines over time, because new and better products become available, which render the old ones obsolete. In exceptional cases, the demand for the use of a program which embodies some ingenious idea may also increase over time, because the use of computers is expanding, and because the existence of the program may become more widely known. But we will confine our analysis to cases in which the social value of a program diminishes over time. It seems intuitive that the more rapidly the value of a program declines, the shorter should the period of protection be that is granted to its producer. In the next section, we will establish the relationship between the optimal period of protection and the useful lifetime of a computer program.

B. A Model to Simulate the Behavior of a Typical Software Firm[1]

The computer software industry is composed of a very large number of firms, some of them controlling a substantial share of the total market, others being very small, sometimes consisting of a single self-employed individual. All of these firms produce products which are similar, but not identical. They are substitutes to some degree, but not perfect substitutes. We can therefore assume that the firms are engaged in monopolistic competition, where each one of them faces a downward sloping demand curve of some sort for its products. Instead of looking at the industry as a whole, with its rather complex structure, it should suffice to analyze the behavior of a single, typical and representative firm to gain insight into how the software industry will tend to react to any proposed policy measure.

---

[1] This section is of a more technical nature and may be skipped without loss by readers who are more interested in policy issues.

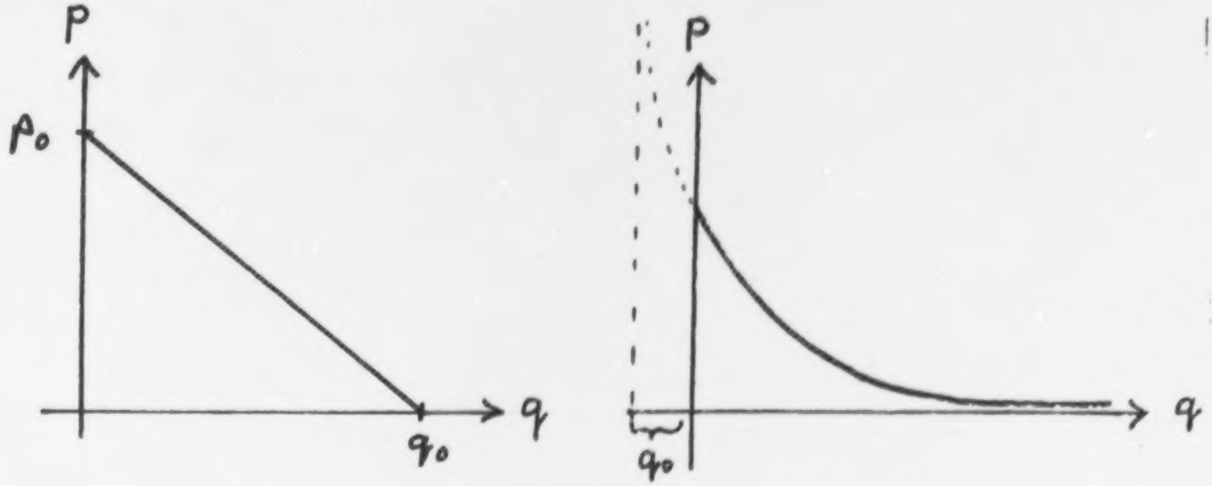
We assume that every firm maximizes the discounted stream of future net revenue. Its behavior (that is, the amount of computer software it produces in each year) is then completely determined by the demand curve it faces as a function of time, its internal cost curve over time, and the nature of protection granted to it. We will specify in turn the assumptions we have made about each of these functions.

In this model we do not distinguish between the sale of a program for unlimited use, and the leasing for a fee to be paid for each use. We simply assume that for any given software package a firm has developed there is some number of users who are willing to purchase it, and that this number is larger, the lower the price is which the firm charges for the package.[1] We also assume that there is a single price for each software package; that is, the firm does not practice price discrimination among users.[1]

---

[1] It has been observed that there have been cases in which a firm was unable to sell a software package at a low price, but found interested buyers when it increased the price, apparently because they believed in the quality of something expensive, but were suspicious about something cheap. But we do not concern ourselves here with such psychological anomalies of decision-makers in a firm.

We first look at the demand curve for one particular software package, during a given relatively short time interval. The simplest shape for the demand curve will be a linear function, and in one version of our model we assume a linear demand curve (Figure 1a). But this may introduce some bias. It seems more natural to us to assume that with a very low or even zero price, the number of users will tend to increase by considerably more than a linear demand curve indicates. We have also experimented with a demand curve which approaches the horizontal axis asymptotically, as shown in Figure 1b.



(a) (b)

Figure 1

(a) A linear demand curve $p = \begin{cases} p_0(1-q/q_0) & \text{if } q \leq q_0 \\ 0 & \text{if } q > q_0 \end{cases}$

(where p is the price and q the corresponding quantity sold

(b) A nonlinear demand curve of the functional form

$p = k(q + q_0)^{-u}$, where $u > 1$.[2]

---

[1] If a firm is allowed to charge discriminatory prices, it can extract a larger amount of surplus from the consumers, and therefore needs a shorter period of protection to be induced to produce a given amount of computer software.

[2] The condition $u > 1$ is necessary because otherwise the consumers' surplus would be unbounded when the price p is equal to zero.

Both of the demand curves shown in Figures 1a and 1b apply to one particular software package, at a given point in time (or, more realistically, for a relatively short time interval, which may, for example, be an entire year). We next define how the demand function declines over time. We assume that the price which any given consumer is willing to pay declines exponentially over time at a constant rate (Figure 2).

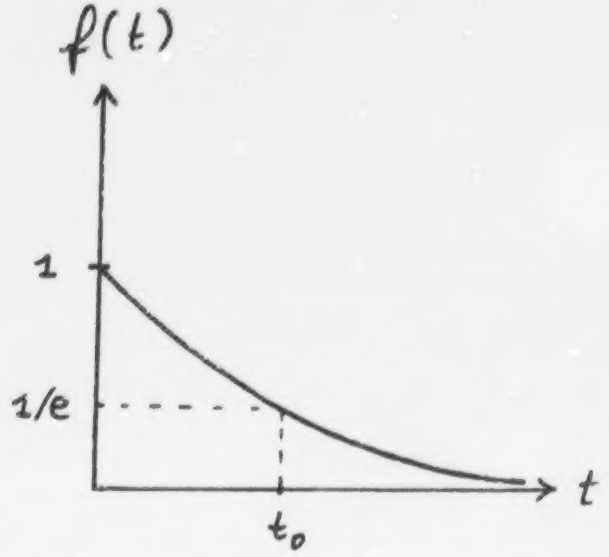


Figure 2

Decline of demand over time. The price each buyer is willing to pay at time t is the price shown in Figure 1 a or b, for t = 0, multiplied by f(t), where $f(t) = e^{-t/t_0}$.

Up to now we have looked at the demand for only one particular software package. (Of course, the same package may be sold over and over again to different users once it is produced, unlike a private good.) We now ask how the various demand curves faced by a firm look for different software packages. It is clear that some of them can be sold at a much higher price than others, but these will also tend to be the ones which are most expensive to produce. We now make the somewhat heroic assumption that there is some measurable unit of output of software, which is defined in such a way that each unit of output can yield to the firm the same amount of revenue (say, e.g., $1000), if the firm is granted a permanent monopoly on the sale of the software it has produced.[1] We also assume that the decline of demand over time progresses at the same rate for each unit of software so defined, so that the demand curve as function of time is identical for each unit of software.[2]

We are now in a position to define a cost function. We assume that each additional unit of software output has an increasing cost, for the two reasons stated. (The firm will have to choose from less and less attractive projects and hire progressively less efficient personnel.) An example of an increasing marginal cost function is shown in Figure 3.

If the firm is granted a permanent monopoly, it will produce the amount X_1 (see Figure 3). If its protection period is limited so that it can earn only the amount MC_0 on each unit of output, then it will produce only the smaller amount X_0 of software, at which marginal cost is equal to marginal revenue per unit.

---

[1] Of course, the shorter the period of protection, the smaller is the revenue the firm can obtain from each unit of output.

[2] Without these simplifying assumptions, the amount of information required to specify the model would simply grow beyond any reasonable bounds.

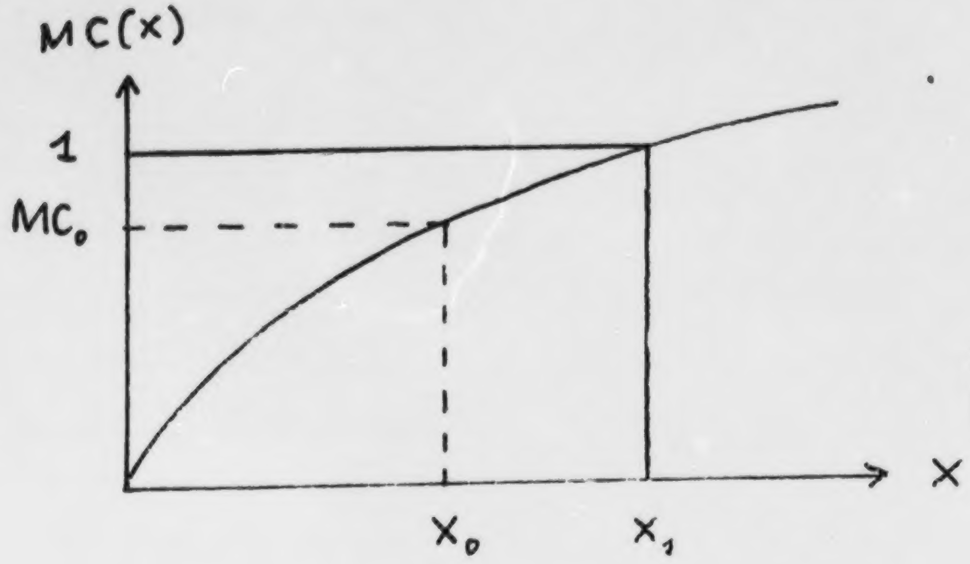


Figure 3

Marginal cost MC(X) as a function of the number of "units" X of software produced: $MC(X) = X^{v}$ with $v > 0$.

We assume that all costs are incurred over a short time period when the program is developed, and that later maintenance costs are negligible. Maintenance costs could easily be built into the model, if information about them were available. But they could simply be discounted back to the point of creation of a new program and would therefore not modify the results obtained here in any essential way.

For the nature of the protection offered we make the simple assumption that the firm is granted a monopoly for a time period of length T after it has produced a new program, and that the program becomes public property thereafter.[1]

To summarize, our model is characterized by the following set of equations:

(1)	$p^{(0)} = p^{(0)}(q)$	price at which a unit of software can be sold q times, per time unit, at time $t = 0$.
(2)	$p^{(t)}(q) = f^{(t)}p^{(0)}(q)$	where $p^{(t)}$ is the price at which a unit can be sold q times per time unit at time t, and $f(t)$ characterizes the decline over time of prices consumers are willing to pay.
(3)	$MC(X)$	marginal cost of producing the Xth unit of output.
(4)	$p^{(t)}(q) = 0$ for $t > T$,	expiration of the monopoly after T time units.

We now want to determine what period T of protection is optimal from society's viewpoint given the profit-maximizing behavior of the firm. The optimal period T will be the one which maximizes the net present value of the discounted stream of social welfare W, which we shall define soon.

Our procedure will be the following. First we will calculate the price the firm will charge for a unit of software during the period of protection, depending on the demand curve. From the duration of the protection period we can then calculate the discounted net present value of the revenue a firm can derive from each unit of computer software. By comparing this revenue with the marginal cost curve MC(X) for production of the X^{th} unit of computer software, we can derive what amount of software the firm will produce during each time period. Next we can compute, for a given period T of protection, the discounted net present value of the future stream of social welfare derived from one unit of computer software. As we have noted earlier, the amount X of computer software produced during one time interval will be an increasing function of the protection period T, whereas the social welfare arising from any particular unit produced will be a declining function of T. By multiplying the amount of software produced by the welfare arising from each unit, we obtain total welfare generated as a function of the period of protection T. This function will increase at first, reach a maximum, and then decline again. The point at which this welfare function reaches its maximum corresponds to the optimal lag period T.

---

[1] One might also consider an institutional arrangement in which firms enjoy a total monopoly at first, with complete freedom of setting prices for their products and then are required gradually to lower the fees at which they make their software packages available to other firms. But it is probably too complicated to administer such a scheme equitably in reality.

We will now execute that program, step by step. First, we will derive the solution in general algebraic terms. Then we will specialize to the particular demand and cost functions shown in Figures 1, 2, and 3.

We will measure social welfare as the sum of consumers' and producers' surplus, as it is customary in economic literature.[1] The producer's surplus or, equivalently, the firm's net profit is the difference between revenue and costs. The consumers' surplus is the difference between what they would be willing to pay at most for a piece of computer software, and what they actually pay. In Figure 4, the consumers' surplus is represented by the area PRP_0. The lower the price, the larger will be the consumers' surplus. Since the producer's costs are independent of the number of times he can sell a given unit of computer software, we look here only at the variable portion of producer's surplus, which alone depends on the quantity sold (or, indirectly, on the

---

[1] This definition of social welfare ignores considerations of distributional equity, and simply adds up the monetary gains to each participant in the economy. For a discussion of other types of welfare functions and their implications for equity, see for example, "The Output Distribution Frontier. Alternatives to Income Taxes and Transfers for Strong Equality Goals," by W.J. Baumol, mimeo, New York University, January 1977.

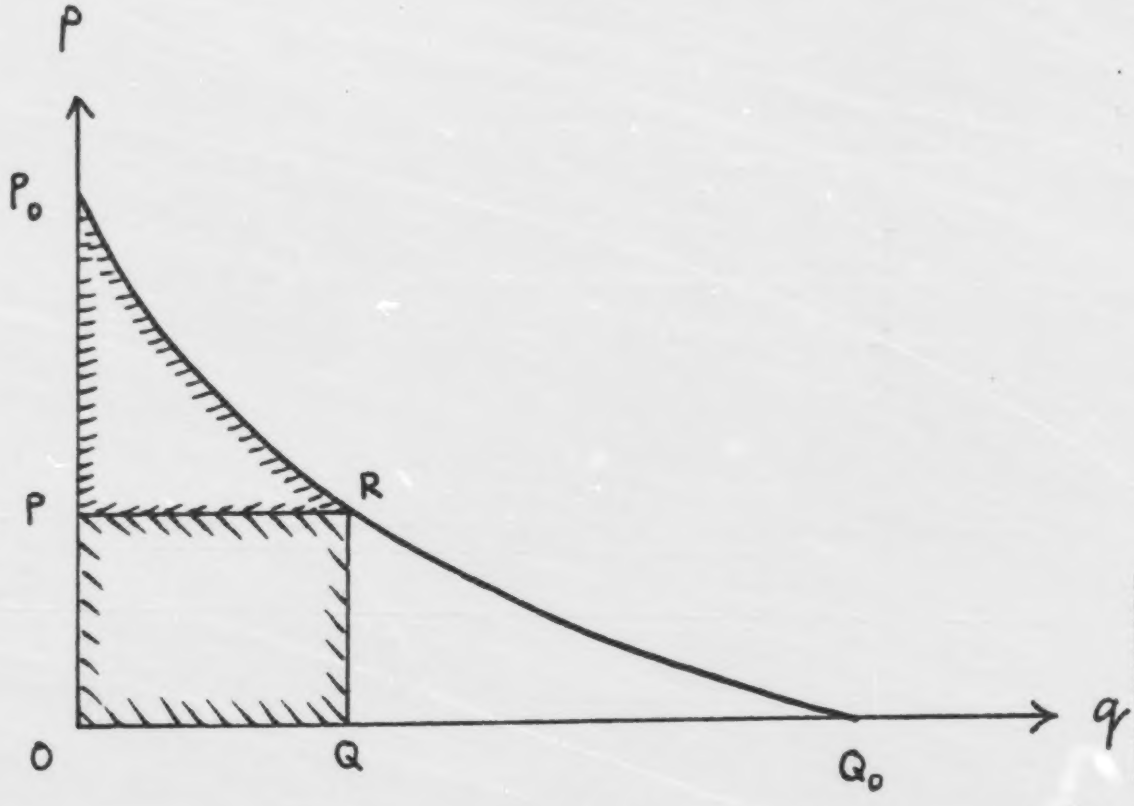


Figure 4

Consumer's surplus PRP_0 and variable portion of producer's surplus OQRP as function of the price OP the firm charges. This figure applies to one particular unit of computer software, during one time interval.

price the firm charges for a unit of computer software). This variable portion of the producer's surplus is represented by the rectangle OQRP in Figure 4.

We can see that as the firm increases the price OP it charges, starting from zero, its revenue OQRP increases at first, but later declines again, as the number of sales it can make gradually drops to zero with increasing price. Somewhere in between, say when OP = p^*, the firm's revenue reaches a maximum. p^* is then the monopoly price the firm will charge as long as it is protected. The corresponding number q of copies sold will be denoted by q^*. After a period of length T, when the firm's copyright expires, the price it can charge drops to zero, and the consumers' surplus then comprises the entire area OQ_0P_0 in Figure 4.

Let us now compute the monopoly price p^*, which the firm will charge per unit of computer software, as a function of the time interval t elaspsed since the production of that unit. Formally using the relationship between price, p, and number of copies sold, q, at any given time t (equation 2) we can state that the monopoly price p^* is given by the value of p which maximizes

$$q \cdot p^{(t)}(q) = qf(t)\, p^{(0)}(q).$$

In other words,

$$q^*(t) \cdot p^{(t)}\,(q^*(t)) \geq q(t) \cdot p^{(t)}(q^{(t)}) \tag{5}$$

for any quantity $0 \leq q(t) < \infty$ sold at time t.

A necessary (but not a sufficient) condition, if $p^{(0)}(q)$ is a smooth function, is that

$$(5') \qquad \frac{d}{dq}\,[p^{(t)}(q)\cdot q] = \frac{d}{dq}\,[p^{(0)}(q)\cdot f(t)\cdot q] = 0 \quad \text{for } q = q^*.$$

We will assume that both the firm and society use a discount rate r. Then the net present value of the revenue a firm can obtain from one unit of software during the period T of protection is equal to

$$(6) \qquad R(T) = \int_0^T p^*(t)q^*(t)e^{-rt}\,dt,$$

where the monopoly price $p^*(t)$ and the corresponding quantity $q^*(t)$ are given by formula (5).

The number of units of software produced, $X(T)$, as a function of the protection period T,can now be obtained from the marginal cost function $MC(X)$ by setting

$$(7) \qquad MC[X(T)] = R(T),$$

and finding the amount $X(T)$ which satisfies (7).

Our next step is to calculate the net present value of the discounted stream of welfare, $W(T)$, generated by one unit of software if the protection period is T.

The contribution to social welfare, $w(t)$, generated by one unit of software, per time unit, at time t during the period of protection is equal to

$$(8) \qquad w(t) = \int_0^{q^*(t)} p^{(t)}(q)dg \qquad \text{for } 0 \leq t \leq T$$

This corresponds to the sum of the two shaded areas $OQRP^0PO$ in Figure 4.

After the protection has elapsed, the social welfare is equal to the consumers' surplus, which now includes the entire area under the demand curve (area $0\ Q_0\ P_0$ in Figure 4):

$$(8') \qquad w'(t) = \int_0^\infty p^{(t)}(q)dq \quad \text{for } t > T.$$

The present value of the discounted stream of the sum of consumers' surplus and the producer's revenue, generated by one unit of software, is then equal to

$$(9) \qquad S(T) = \int_0^T w(t)e^{-rt}dt + \int_T^\infty w'(t)e^{-rt}dt.$$

Since $w'(t) > w(t)$, (the welfare contribution of a unit of software is larger without protection than with protection), we see that $S(T)$ reaches its maximum at $T = 0$ and declines with increasing T.
To obtain the contribution of a unit of software to social welfare, we would have to subtract its production costs. But average production costs will depend on the amount X of software produced, and we will therefore have to take each into account separately. The cost of producing $X(T)$ units of software, $C(T)$, is equal to

$$(10) \qquad C(T) = \int_0^{X(T)} MC(x)dx$$

No discounting is necessary here, since we assumed that all costs are incurred at the point of production when $t = 0$.

The contribution to social welfare of all units of software produced during one time unit at $t = 0$, discounted to infinity, as a function of the protection period T, is now equal to the amount $X(T)$ of computer software produced, multiplied by the consumers' and producer's surplus, $S(T)$, generated by each unit of software, minus the production costs $C(T)$:

$$W(T) = X(T) \cdot S(T) - C(T). \tag{11}$$

Figure 5 shows some typical shapes for $X(T)$, $S(T)$, $C(T)$ and $W(T)$. The value of T which maximizes $W(T)$, T^*, is the optimal period of protection:

$$W(T^*) \geq W(T) \quad \text{for } 0 \leq T < \infty \ . \tag{12}$$

Results obtained for the optimal period of protection as a function of various parameters, such as the discount rate, the average useful life of a software package, and the structure of costs and demand will be presented in the next subsection. We will also compare the maximum amount of welfare that can be obtained through a copyright system, when the optimal protection period is granted, with the social welfare that could be achieved in principle through some form of central planning. In the latter case, we will assume that a firm produces the socially optimal amount[1] of computer software, financed through gevernment support, and makes the software it has produced available for free (that is, at marginal costs) to anyone who wants to use it.

---

[1] For the socially optimal amount to be produced, the marginal cost of the last unit of software produced must be equal to marginal social benefit.

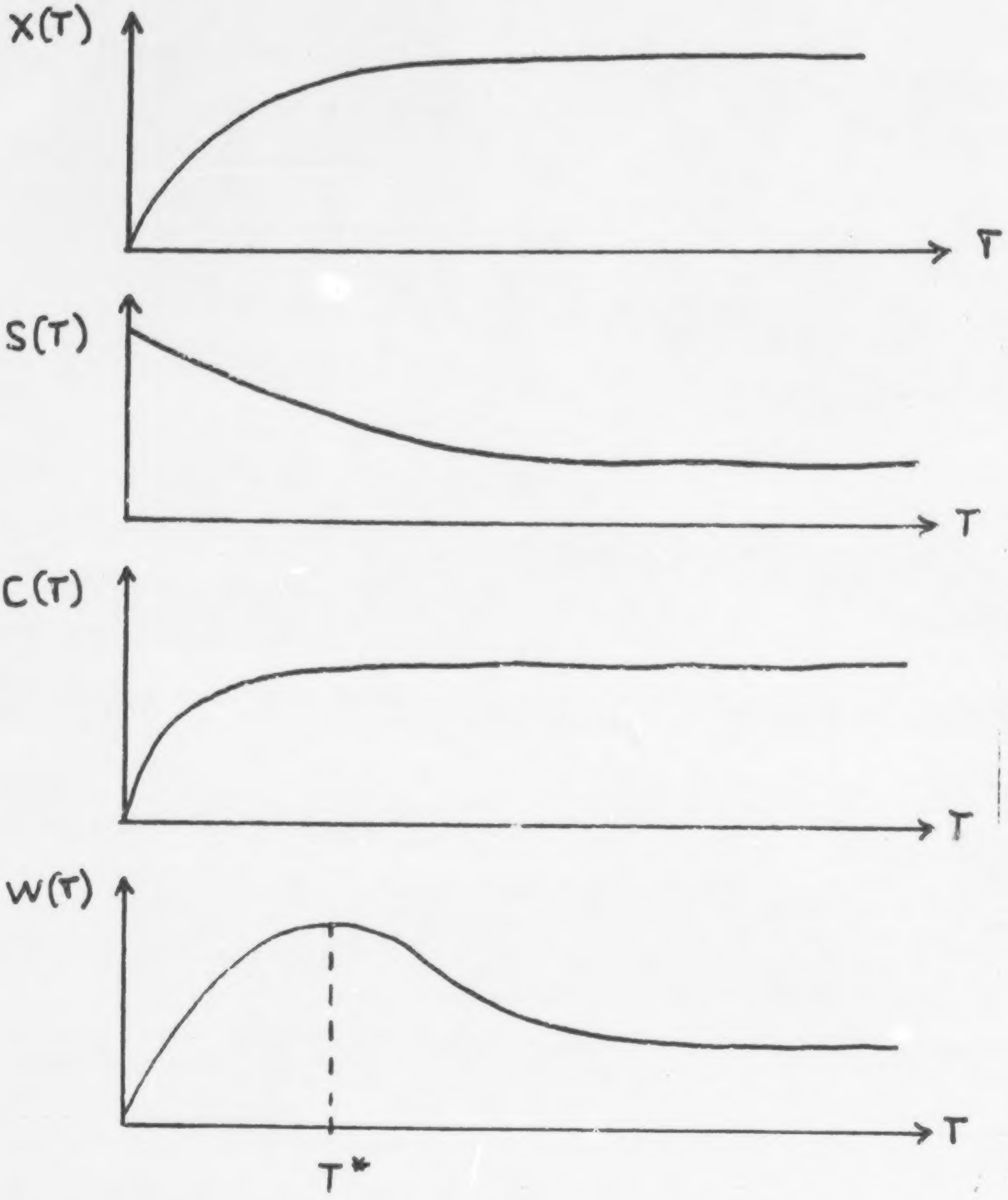


Figure 5

Amount of software produced, X(T), consumers' surplus and producer's revenue generated per unit of software, S(T), total cost of all units of software produced, C(T), and present value of discounted stream of welfare, W(T), as functions of the protection period T. T* is the optimal period of protection.

C. Results Obtained from the Model and Their Policy Implications

Let us first consider a linear demand curve (as shown in Figure 1a). As is shown in Appendix IV-1, the optimal period of protection, T^*, can be expressed by the equation

$$T^* = \frac{1}{r+1/t_0} \ln \frac{v+3}{v-1} \quad (13)$$

where

r is the discount rate

t_0 is the average useful life of a unit of computer software (the willingness to pay decays proportional to e^{-t/t_0})

v is a measure for the degree of diseconomies to scale in producing computer software. (Marginal costs are proportional to X^v, where X is the amount of computer software produced by the firm.)

From this formula it is apparent that, at least under the assumptions made here, the optimal protection period T^* has the following relationship to the parameters on which it depends:

Principle 4.1 If the discount rate, r, increases, then the optimal period of protection T^* decreases, provided other parameters remain unchanged.

This principle implies that the faster we discount future income, i.e., the less we value future income compared to present income, the shorter should the period of protection be. With a high discount rate, society is less interested in how much software is being produced, but is more eager to obtain the full benefit of free use from whatever is being produced as soon as possible. On the other hand, if the social discount rate is low, society is willing to grant a longer protection period and make some sacrifices in the near future in order to promote the production of software to a fuller extent and have more to live on in the more distant future. This relationship corresponds to what one would naturally expect.

Principle 4.2 If the average useful lifetime, t_0, of a unit of computer software increases, the optimal period of protection should be increased.

This principle, again, corresponds to our intuition that the longer it will take until a piece of computer software is superseded by a better one, the longer it is worth to protect it, in order to have a greater amount of computer software produced.

Principle 4.3 The stronger the diseconomies of scale are in the production of computer software (as measured by the exponent v), the shorter ought to be the optimal period of protection T*.

This principle is somewhat less obvious than the first two principles. But it too has a natural interpretation in hindsight. If there are very strong diseconomies of scale, this means that there is some limited number of software projects which it is worthwhile to undertake, but anything beyond that selection will be very expensive to produce, as costs increase rapidly. In this situation, a longer protection period, which is designed to give more revenue per unit of software to the producer and in this way to encourage him to produce more, will have very little effect. Only a few projects, if any at all, will be made cost-effective if they were unprofitable under a shorter protection period.

That means, by increasing the protection period, society would forego a large amount in potential welfare from the free use of existing computer software, in order to stimulate the production of only a small additional amount of computer software. A short period of protection is therefore indicated in the presence of strong diseconomies of scale. Just the opposite is true if there are only weak diseconomies of scale or almost constant returns to scale. In that case, a slight increase in revenue per unit of software produced will now make a wide range of new projects financially feasible - projects which were unprofitable under a shorter protection period. Therefore, by giving up only a limited amount of benefit from existing computer software through a prolongation of the producer's monopoly, society will induce producers to create a considerably larger amount of software. Therefore, if diseconomies of scale are weak, a long protection period is indicated.[1]

We turn next to the nonlinear demand curve shown in Figure 1b. Here, monopoly pricing excludes a much larger number of potential users who are willing to pay only a low price for the use of a software package than was the case when the demand curve is linear. We will therefore expect the differences in social welfare between monopoly pricing during the protection period and free use after the lapse of the protection period to be larger than in the case of a linear demand curve. In other words, the penalty paid by society during the protection period is higher than in the previously considered case.

---

[1] Mathematically oriented readers will have noticed that the optimal protection period under the model assumptions made here is finite only if $v > 1$, that is, marginal costs of production increase at a faster rate than the amount produced. This implies that total costs increase more rapidly than the square of output. If the diseconomies of scale are any weaker, then the optimal protection period is infinite.

As is shown in Appendix IV-2, the optimal period of protection, T^*, in the case of the nonlinear demand function assumed is

$$(14) \qquad T^* = \frac{1}{r+1/t_0} \ \ell n \ \frac{1}{1-z^*} \quad \text{with } z^* = \left(\frac{u}{u-1}\right)^u \cdot \frac{u-1}{uv+2u-1}$$

where

r is the discount rate

t_0 the average useful life at a unit of software

v is a measure for the degree of diseconomies of scale, and

u indicates how rapidly the willingness to-pay declines with increasing numbers of consumers ($p = k(q_0+q)^{-u}$, where q is the number of copies of a unit of software sold at price p, and k and q_0 are parameters defined earlier.) The higher u is the less demand responds to a change in price, i.e., the less price elastic is demand.

As before, (in equation 13) we find that the optimal period of protection decreases with increasing discount rate r, increases with increasing average life time t_0, and decreases with increasing diseconomies of scale in production, represented by v. In addition, we have the new relationship

Principle 4.4 The less demand responds to a change in price, the longer should the period of protection be.[1]

This principle agrees with the intuitive relationship between the shape of the demand curve and the penalty paid by society for having a monopoly, which we just discussed. If the willingness to pay declines rapidly for additional potential users, i.e., if demand is inelastic, then little is sacrificed by having a non-zero monopoly price charged for the use of software, and society can afford to grant a longer period of protection. On the other hand, if monopoly pricing excludes many potential users from taking advantage of existing software - that is, if demand is more price-elastic - then the loss resulting from a monopoly is more serious and a shorter period of protection is appropriate.

---

[1] It is a little tricky to see that T^* increases with u (by the definition of u, the higher the value of u, the more rapidly the demand curve declines). But we will soon see graphs which show the dependency of T^* on u and other parameters.

D. Data Requirements and Estimates of Optimal Protection Periods

In order to estimate the shape of the demand curve for computer software, it would be necessary to collect information on the number of sales of a given software package that a firm has been able to make at various prices. If demand can be represented sufficiently well by a linear demand curve, then this information alone is sufficient to permit the calculation of an optimal period of protection from the discount rate, the average life-time of software, and the cost curve. It is not necessary to have any more detailed estimates of demand.[1] How the discount rate r is to be estimated should pose no problem. The average life-time of a piece of software, t_0, can be obtained from information on how rapidly the sales-price, or the number of sales of a software package at a constant price, decline. To estimate the degree of diseconomies of scale in the production of software (as represented by the parameter v), it would be necessary to ask software firms by how many percent their costs would increase if they were to increase their output by, say, 10 percent (where output is measured by the dollar volume of sales). Given these three parameters, r, t_0 and v, the optimal period of protection T* can be calculated from equation (13). For example, if we take a 10 percent annual discount rate (r = .10), a 5-year average life-time of a software package, and a parameter v = 2 (indicating that marginal costs are proportional to the square of the quantity of software produced), then we obtain

---

[1] That is, it is not necessary to know the slope of the demand curve, and the intercepts p_0 and q_0 in Figure 1a.

$$T^* = \frac{1}{r+1/t_0} \ln \frac{v+3}{v-1} = \frac{1}{.10+.20} \ln \frac{5}{1} = 3.33(1.61) = 5.36,$$

i.e., a copyright protection for a period of about 5 years is optimal.

Figures 6 to 8 indicate how the optimal protection period T* depends on r, t_0 and v.

How serious is it if the length of the protection period departs from the optimum? To gain insight into this problem, we can calculate the relative efficiency of various protection periods T, where the relative efficiency indicates what fraction of the maximum feasible amount of welfare[1] is reached by a copyright of a duration of T years. As is shown in Appendix IV-1, for a linear demand curve the relative efficiency, E(T), is given by

$$(15) \qquad E(T) = (1/v)\ (z/2)^{1/v}\ (v+1 - (v+3)z/4),$$

where $z = 1-e^{-sT}$ and $s = r + 1/t_0$.

Figure 9 shows the relative efficiency E(T) as a function of the protection period T, for the set of parameters used in the example

---

[1] This maximum could be reached through some ideal form of central planning.

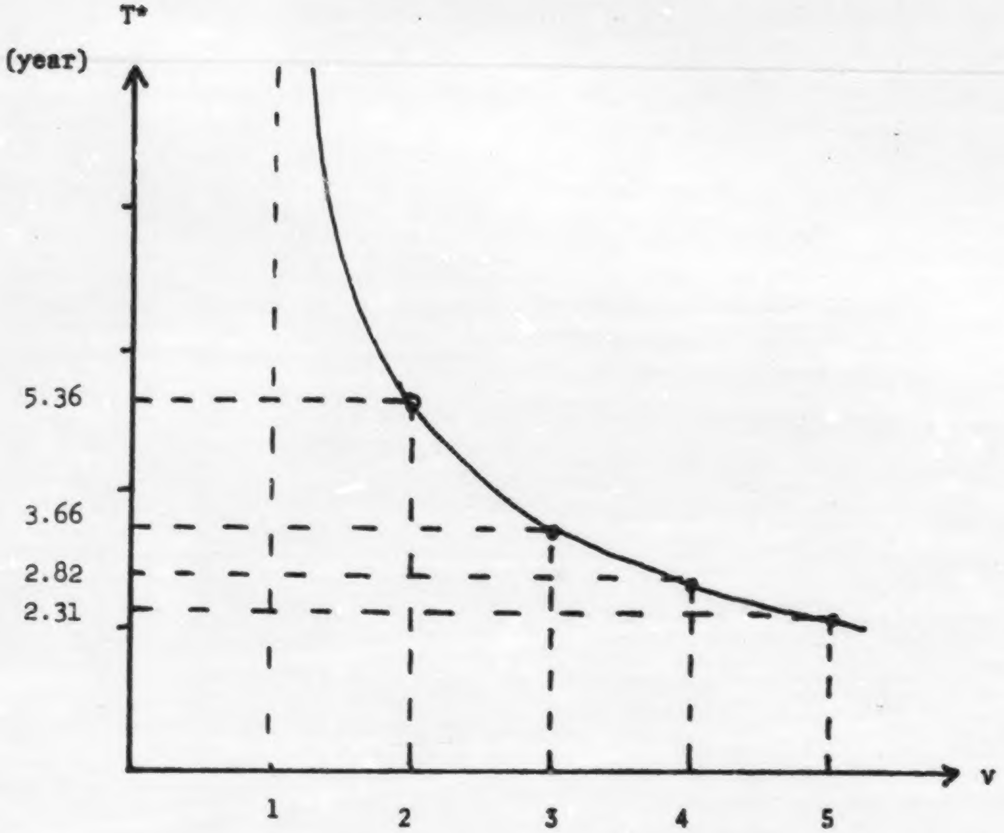


Figure 6

The optimal period of protection, T*, as a function of the degree of diseconomies of scale v, for a discount rate r = 10 percent per year and an average life-time for computer software of t_0 = 5 years, if the demand curve is linear.

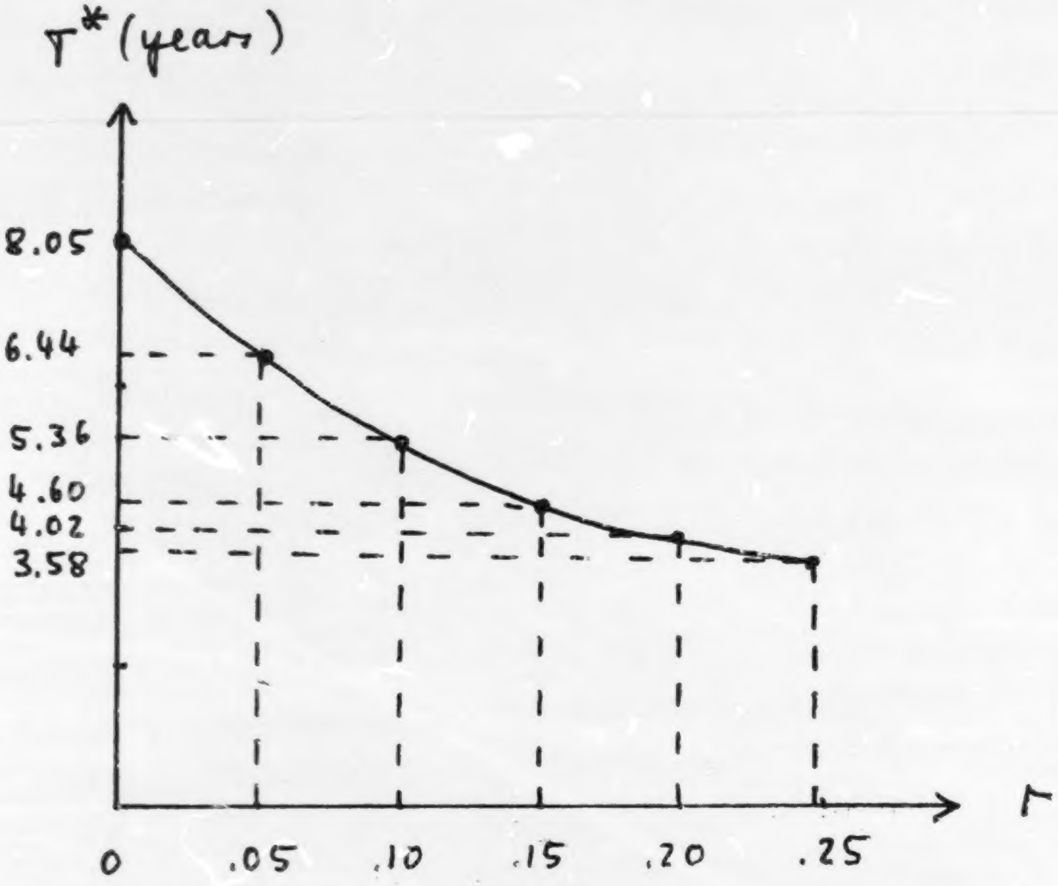


Figure 7

The optimal period of protection, T*, as a function of the annual discount rate r, for an average life-time for computer software of $t_0 = 5$ years, and diseconomies of scale of $v = 2$. (Case of a linear demand curve).

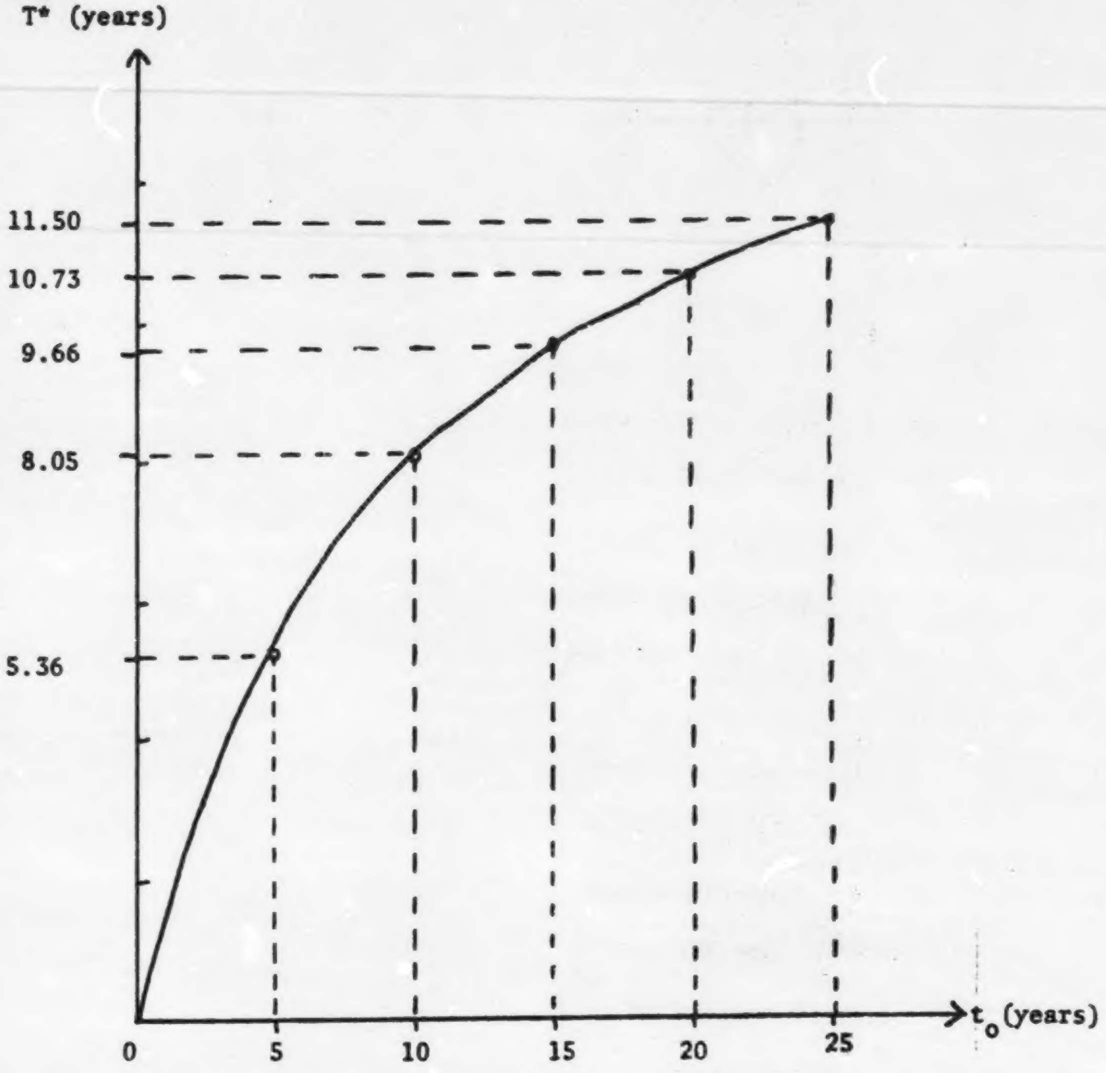


Figure 8

The optimal period of protection, T*, as a function of the average life-time for computer software, t_0, for an annual discount rate r = .10 and diseconomies of scale in production given by v = 2. (Case of a linear demand curve.)

just given.[1] Table 1 lists efficiencies for some selected values of T.

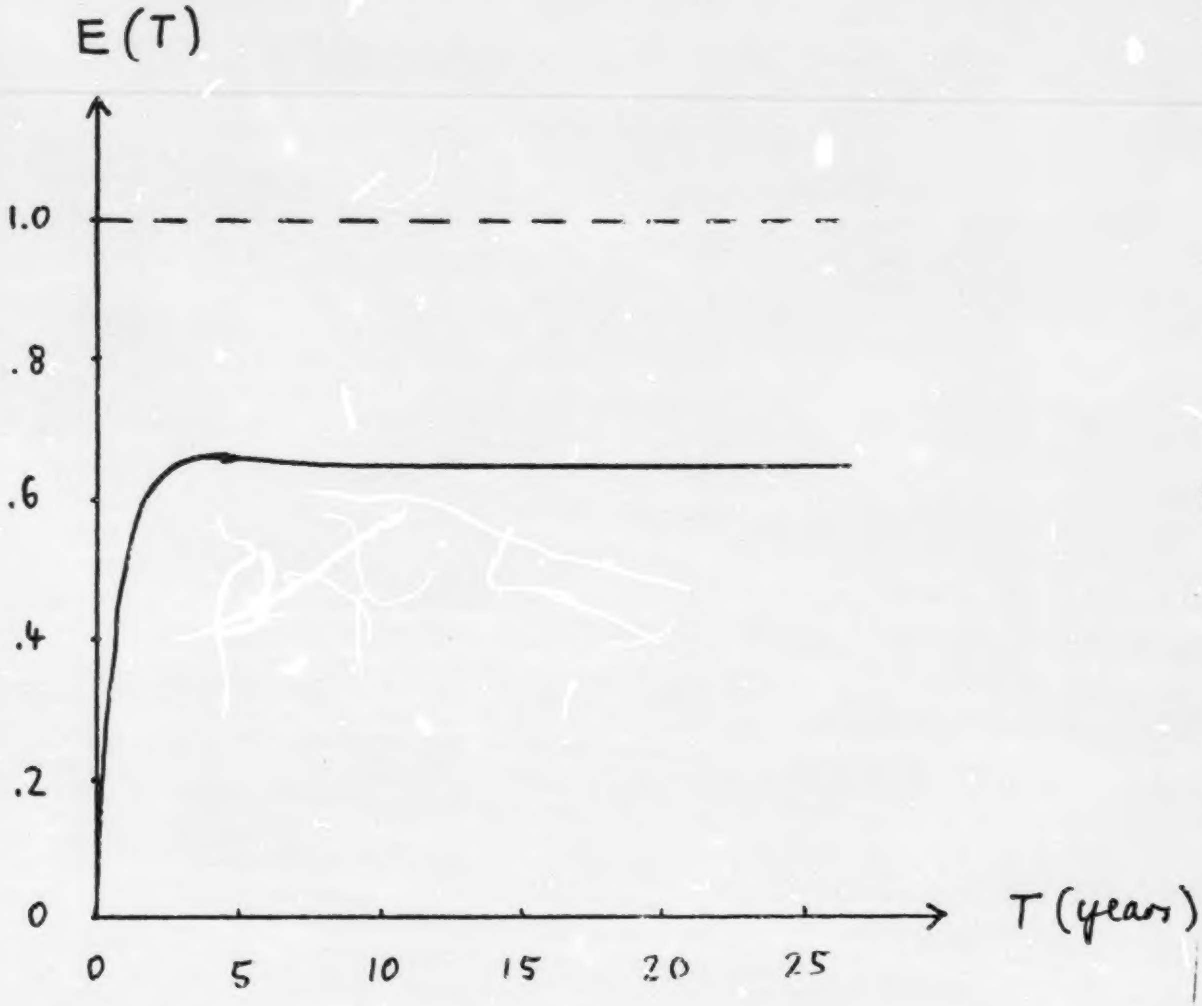


Figure 9

Relative efficiency E(T) as a function of the protection period T, for an average life-time of software of $t_0 = 5$ years, an annual discount rate $r = 0.10$, and diseconomies of scale measured by $v = 2$, where marginal costs are equal to the v'th power of output x. (Case of a linear demand curve.)

---

[1] This diagram corresponds to the social welfare W(T) shown in Figure 5.

T	0	0.1	.25	.5	1	2	5.36	10	1000
E(T)	0	.180	.277	.373	.482	.579	.632	.625	.619

Table 1

Relative efficiency E(T) as a function of the protection period T (in years), for the same parameters as in Figure 9 (for a linear demand curve).

As Figure 9 and Table 1 indicate, the maximum of the efficiency is very flat. There may be some loss in efficiency if the protection period T is much too short, but even if T is only 2 years (less than half of the optimum of 5.36 years), the loss in relative efficiency is less than 10 percent. And if the protection is granted for as long as 1000 years, the loss in efficiency is only about 2 percent.

The fact that the relative efficiency curve is almost flat to the right of the optimum was also pointed out by Nordhaus, in a slightly different context.[1] This is typical for linear demand curves. As we shall see soon, however, for some non-linear demand curves, too long a protection period can bring a considerable loss in welfare.

Table 2 lists the optimal protection period T for various parameter combinations. It also shows that relative efficiency that can be obtained with the optimal protection period, and the relative efficiency that would result from permanent protection (when T approaches infinity).

---

[1] See Invention, Growth and Welfare by W.D. Nordhaus (MIT Press, Cambridge, MA, 1969) esp. pp. 83-84. In his model, Nordhaus considers patent protection on cost-saving inventions.

```
                V= 0.1   V= 0.2   V= 0.5   V= 1.0   V= 2.0   V= 5.0   V=10.0

L=   1   T◆    ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆    1.463    0.630    0.334
         E◆     0.003    0.062    0.312    0.500    0.632    0.758    0.829
         E      0.003    0.063    0.313    0.500    0.619    0.696    0.723

L=   2   T◆    ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆    2.682    1.155    0.613
         E◆     0.003    0.062    0.312    0.500    0.632    0.758    0.829
         E      0.003    0.063    0.313    0.500    0.619    0.696    0.723

L=   5   T◆    ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆    5.365    2.310    1.226
         E◆     0.003    0.062    0.312    0.500    0.632    0.758    0.829
         E      0.003    0.063    0.313    0.500    0.619    0.696    0.723

L=  10   T◆    ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆    8.047    3.466    1.839
         E◆     0.003    0.062    0.312    0.500    0.632    0.758    0.829
         E      0.003    0.063    0.313    0.500    0.619    0.696    0.723

L=  20   T◆    ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   10.730    4.621    2.451
         E◆     0.003    0.062    0.312    0.500    0.632    0.758    0.829
         E      0.003    0.063    0.313    0.500    0.619    0.696    0.723

L=  50   T◆    ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   13.412    5.776    3.064
         E◆     0.003    0.062    0.312    0.500    0.632    0.758    0.829
         E      0.003    0.063    0.313    0.500    0.619    0.696    0.723

L=100    T◆    ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   ◆◆◆◆◆◆   14.631    6.301    3.343
         E◆     0.003    0.062    0.312    0.500    0.632    0.758    0.829
         E      0.003    0.063    0.313    0.500    0.619    0.696    0.723

◆◆◆◆◆◆ INDICATES THAT THE OPTIMAL PROTECTION PERIOD IS INFINITE
```

Table 2

Optimal protection period (T*), optimal relative efficiency (E*) and relative efficiency obtained with permanent protection (E), for a <u>linear demand curve</u>, for various values of v, the degree of diseconomies of scale in software production, and various average life times L (in years). The annual discount rate r is assumed to be 10 percent.

As Table 2 shows, the highest relative efficiencies which can be obtained through a copyright system are achievable when there are very strong diseconomies of scale in the production of software, i.e., when the parameter v is high. A high value of v means that the protection period should be kept relatively low, and that additional protection, which increases a firm's revenue, will induce it to increase its production only by very little. If $v = 10$, that is, if a 10 percent increase in output raises marginal costs by a factor of $(1.1)^{10} = 2.6$, then a copyright of optimal duration will yield about 83 percent of the maximum amount of welfare that is achievable in principle. But if diseconomies of scale are not as strong, then the efficiency of a copyright system, even one of optimal duration, is considerably lower. If marginal costs are proportional to output (i.e., $v = 1$), then the efficiency is 50 percent. If v is as low as 0.1, meaning that a 10 percent increase in output raises marginal costs only by 1 percent, then the efficiency of a copyright system is less than 1 percent of what is feasible in principle, with the right amount of public funding of software development.

Let us now look at some results for a non-linear demand curve for software, of the type shown in Figure 1b:

$$p(q) = k(q+q_0)^{-u}$$

As is shown in Appendix IV-2, the optimal period of protection in that case is given by

$$(14) \quad q\, T^* = -\frac{1}{s} \ln (1-z^*)$$

where $z^* = \left(\frac{u}{u-1}\right)^u \frac{u-1}{uv+2u-1}$ and $s = r+1/t_0$.

Here the parameter u indicates how rapidly the price drops with an increase in supply. For example, if $u = 2$ and $v = 2$, and $t_0 = 5$ and $r = .10$ as before, we find $s = 0.3$, $z^* = 4/7$, and $T^* = 2.82$. The behavior of the optimal protection period with respect to the parameters v (diseconomies of scale), r (discount rate) and t_0 (average life-time of a software package) is similar as before in the linear case. What is new is the way T^* depends on u, where a low value of u indicates more elastic demand, and a high value of u indicates inelastic demand for software. Figure 10 shows how T^* increases with increasing values of u, given that all the other parameters are held fixed.

Again, we ask how serious it is to depart from the optimum duration for the protection period. In Appendix IV-2 it is shown that the relative efficiency, E(T), of a protection period of length T is given by

$$(16) \qquad E(T) = \frac{v+1}{v} (yz)^{1/v} \left(1 - \frac{u}{u-1} yz\right) - \frac{1}{v} (yz)^{(v+1)/v}$$

with $z = 1-e^{-sT}$, $s = r+1/t_0$, and $y = (u/(u-1))^{-u}$.

In Figure 11, a graph of the function E(T) is shown for a specific combination of parameter values. Table 3 lists selected values of E(T).

As Figure 11 and Table 3 clearly show, there may be a substantial penalty for granting too long a protection period, if demand for software is not linear and rather elastic in the lower right tail of the

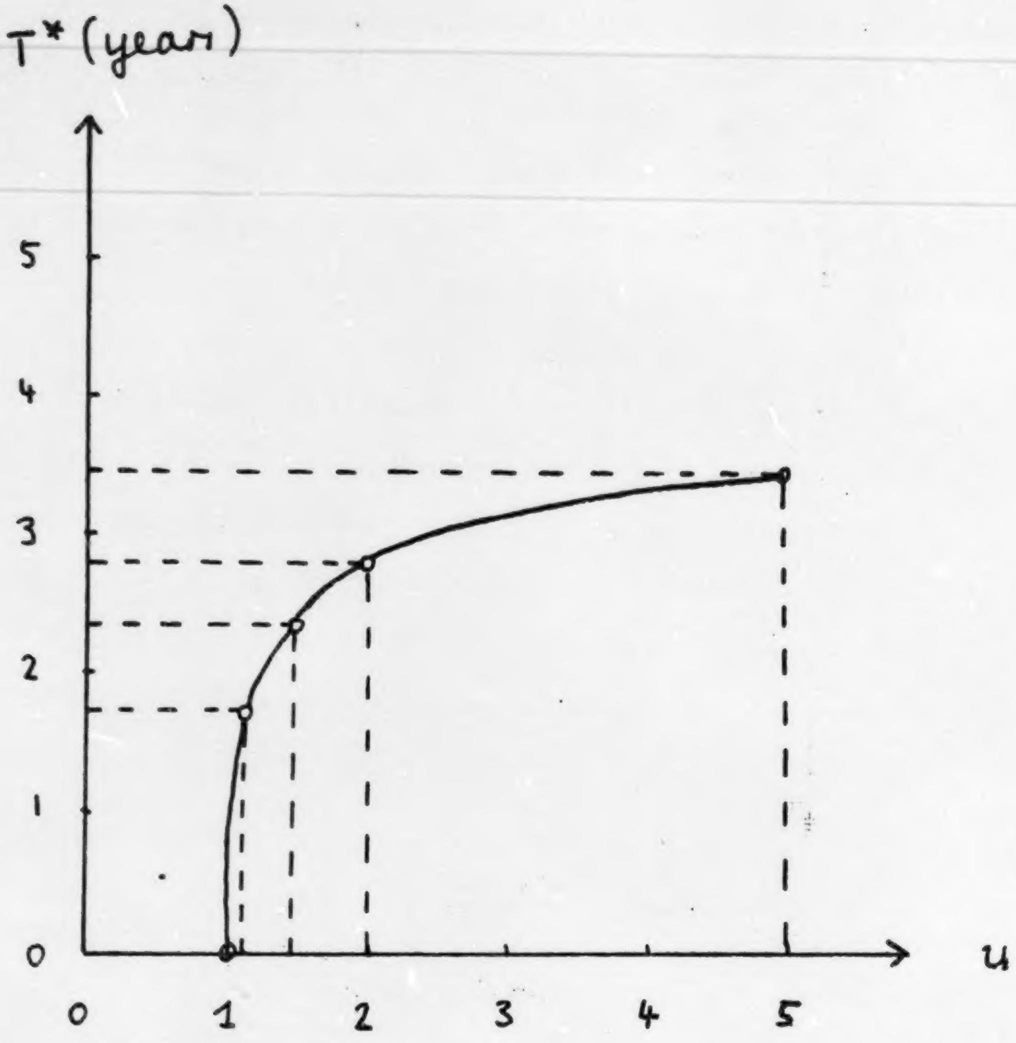


Figure 10

Optimal protection period (T*) as a function of u, where high u means inelastic demand. Parameter values are v = 2 (diseconomies of scale), t_0 = 5 (average life-time of software in years) and r = .10 (annual discount rate). The demand curve is nonlinear, as shown in Figure 1b.

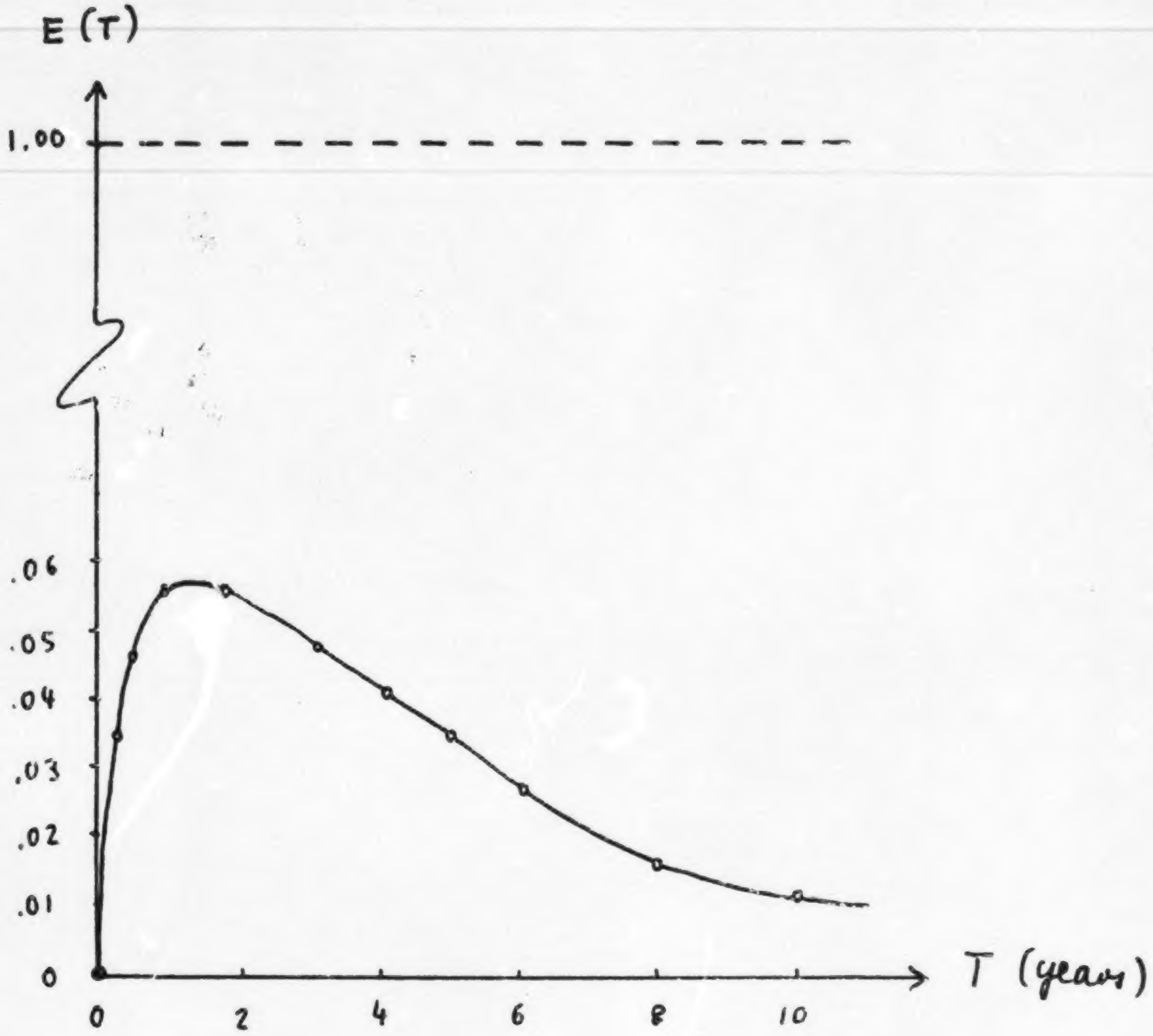


Figure 11

Relative efficiency E(T) as a function of the protection period T, for u = 1.01 (indicating a demand elasticity close to unity), v = 2 (diseconomies of scale), t_0 = 5 (average life-time of a software package, in years) and r = .10 (discount rate). (Case of a nonlinear demand curve.)

T	0	.01	.05	.10	.2	.5	1.0	1.425
E(T)	0	.008	.018	.024	.003	.047	.056	.057

T	2	3	4	5	6	8	10	20	∞
E(T)	.056	.048	.040	.033	.027	.018	.013	.006	.006

Table 3

Relative efficiency E(T) for selected values of T, for the same parameters as in Figure 11 (for a nonlinear demand curve).

demand curve. The reason is that during the protection period a very large number of potential users are excluded from enjoying the benefits of already existing software, because of the monopoly price being charged for it. Thus, any extension of protection, beyond the mininum period needed to encourage the production of at least some software, will necessarily reduce public welfare by a large amount.

Table 4 shows the relative efficiency of a copyright system for various combinations of the parameters u and v.

What is significant about the results shown in Table 4 is that even for some rather plausible values of u and v, the relative efficiency of a copyright system of optimal duration is very low, often less than 1 percent. Only when diseconomies of scale are very strong (i.e., v is high) does a copyright system come close in efficiency to a publicly financed software industry, because in that case very little additional output can be obtained from an increase in investments. But in other

		V= 0.1	V= 0.2	V= 0.5	V= 1.0	V= 2.0	V= 5.0	V=10.0
U= 1.01	T*	9.581	6.689	3.943	2.454	1.425	0.638	0.333
	E*	0.000	0.000	0.000	0.005	0.057	0.278	0.496
	E	0.000	0.000	0.000	0.001	0.006	0.021	0.031
U= 1.10	T*	******	13.904	5.347	3.120	1.766	0.780	0.406
	E*	0.000	0.000	0.003	0.043	0.171	0.431	0.619
	E	0.000	0.000	0.003	0.025	0.076	0.143	0.175
U= 1.50	T*	******	******	9.653	4.520	2.444	1.065	0.553
	E*	0.000	0.000	0.033	0.143	0.316	0.555	0.703
	E	0.000	0.000	0.033	0.126	0.236	0.337	0.378
U= 2.00	T*	******	******	******	5.365	2.824	1.226	0.637
	E*	0.000	0.002	0.062	0.200	0.378	0.599	0.731
	E	0.000	0.002	0.063	0.188	0.313	0.417	0.457
U= 5.00	T*	******	******	******	6.851	3.429	1.483	0.773
	E*	0.000	0.007	0.120	0.286	0.459	0.652	0.764
	E	0.000	0.007	0.120	0.280	0.413	0.514	0.552
U= 10.00	T*	******	******	******	7.359	3.614	1.562	0.815
	E*	0.000	0.010	0.139	0.310	0.480	0.665	0.772
	E	0.000	0.010	0.139	0.306	0.440	0.539	0.575
U=100.00	T*	******	******	******	7.829	3.775	1.631	0.852
	E*	0.000	0.013	0.155	0.331	0.498	0.676	0.779
	E	0.000	0.013	0.155	0.327	0.461	0.559	0.594

****** INDICATES THAT THE OPTIMAL PROTECTION PERIOD IS INFINITE

Table 4

Optimal protection period (T*), optimal relative efficiency (E*) and relative efficiency obtained with permanent protection (E), for various values of u (when a high u indicates inelastic demand) and v (high v indicates strong diseconomies of scale in production). The average life-time of a copyright, t_0, is assumed to be 5 years, and the discount rate r = .10. (Case of a nonlinear demand curve.)

cases, where costs do not increase prohibitively with an expansion of output, it turns out that a reliance on profit motive of private firms does not help to generate an amount of computer software close to the social optimum. Even when these firms are given just the proper amount of property right protection, not too little and not too much, it will still be observed that many potential investments in the development of computer software, which are worthwhile to be undertaken from the point of view of society as a whole, will be unprofitable for any one firm. The deeper reason for this state of affairs is, of course, the public goods aspect of information, which can be shared without being used up.

These findings do not imply that a copyright system for computer software is a bad idea, and should be abandoned in favor of the current reliance on trade secret laws. For reasons expounded elsewhere in this report, trade secrets are even worse from the point of view of public welfare. What these results mean is that if the software industry has a certain cost and demand structure (moderate diseconomies of scale and elastic demand), then a reliance on private enterprise to produce a desirable amount of computer software will exact a high price from society. Whether such demand and cost conditions prevail in the computer software industry has to be determined from empirical observation. We hope that this study has indicated how important it is to undertake such an empirical investigation.

E. Protection vs. Public Subsidies for the Production of Software

As we have seen at the end of the previous section, in some cases the granting of a monopoly to producers of computer software for a period of optimal length can yield a contribution to public welfare which comes close to what would be possible in principle through government planning. But in some other cases, especially when demand for computer software is quite elastic, the welfare loss caused by protection of 1000 years reduces welfare only very slightly over a patent of optimal duration, which we found to lie between 6 and 20 years, depending on the parameter assumptions. But as we have seen, these findings critically depend on the model assumptions made.

If it should turn out from empirical observations that private firms do not produce the optimal amount of computer software, because many projects that would be worthwhile to undertake from the viewpoint of society as a whole are not profitable to an individual firm, even if it is granted a monopoly of limited duration, then it is worth it to consider whether alternatives to statutory protection should be explored. Among other possibilities we think primarily of increased public support for the production of computer software, at universities and other non-profit research organizations.

If the development of new computer software is supported with public funds, there is no need to grant a monopoly of a certain duration to the producer, and new computer programs and data bases can be made available immediately to all potential users and benefit the entire economy. Of course, public funds must come from somewhere, and they will have to be financed through taxes imposed by the government. But taxes can be spread among a large group of firms and other users, scaled according to each one's ability to pay, and they do not discourage potential users of software to the same extent as a price charged under a monopoly.

In a related study, two of the coauthors of this report[1] have shown that in a simple economic model with innovation the economy grows faster if innovation is sponsored by the government and the results are made available to everybody, than if the government grants a patent of optimal duration to innovators and leaves innovation to private firms. In the system with a patent, less will be invested in innovation than is socially optimal. It was also found that the loss in welfare resulting from the misallocation of resources under a temporary monopoly is considerably larger than the monopoly profits earned by the innovating firm.

In the model studied in Baumol and Fischer, innovation played a central role in the entire economy, and a short-fall in the amount of innovation forthcoming led to a retardation of the growth of the whole economy. Even a slight reduction in the annual growth rate will ultimately make a great difference in the output level of the economy, if the growth rate is compounded year after year over an extended period.

If we look specifically at computer software, and not innovations in general, then the effect on the U.S. economy of underproduction in that sector will be less serious, at least for the present time, since the software industry amounts to less than one percent of all economic activities. We would therefore not expect a noticeable slowdown in economic growth to result, even if there is some measurable loss of welfare because less software is being produced than would be socially desireable. But this situation could change if our economy were to rely more and more on computers and automation. And trends indicate that this will indeed soon be the case. We shall return to this point in the final paragraphs of this section.

---

[1] See "Optimal Lags in a Schumpeterian Innovation Process" by W.J. Baumol and D.M. Fischer, mimio, New York University, 1976.

Of course, a system of publicly financed production of computer software is not without problems of its own. Having the government pay for the production of computer software and make it available for free to everybody on request eliminates the test of the market, whether a newly produced software package meets a real demand. Once this test for profitability is eliminated, it is necessary to replace it by equivalent, or if possible better criteria to evaluate which activities in the software field deserve to be supported. Without any screening, researchers will have a tendency to do what is interesting rather than what is needed for economic progress.

What type of activities in the software field should be supported with public funds? As was discussed in Chapter III-B, firms who develop software which is very similar to what already exists, simply to capture a share of a lucrative market for themselves without contributing to an increase in public welfare, will not be able to justify such activities for public support. But other institutions do deserve to receive government support in the public interest. Among them will be especially those who are engaged in highly innovative activities which hold potentially great promises for uncovering new opportunities. These activities are more risky, and their benefits may be realized only in the more distant future, making them commercially less attractive. Among the type of projects primarily deserving public support are the development of new computer algorithms; the mathematical specification of algorithms in unambiguous form; the development of translating systems which allow programs to be written in machine independent form, without reference to the specific hardware used; the development of more flexible programming languages, which are free of artificial restrictions stemming from historical developments; the automation of the programming process itself, allowing a user to specify only what would be needed to explain to an intelligent person what a program is designed to do, and leaving all the details of memory allocation, efficient coding, etc. to a powerful compiler; and, finally, research in artificial intelligence, which may one day make feasible new types of thought processes which we would never have imagined.

In the absence of a commercial test of what is demanded, there ought to be a constant dialogue between users and developers of new software, to assure that the needs of users are met.

An alternative to either statutory protection or government support is the reliance on trade secrecy to give producers of software sufficient financial incentives, which is the current situation. Trade secret laws, which allow firms to conclude binding agreements among each other about conditions for the sale and lease of computer software, imply that there is some period of protection of a random duration and of decreasing efficiency, as the number of imitations increases. The secret may be lost through unauthorized users, for example an employee who changes his job and keeps a copy of a computer program developed by his previous company.

The resulting random period of protection will hardly happen to be of optimal length. But what is even more serious, the lack of public disclosure of computer programs, out of fear of unauthorized use, reduces the revenue a producer might win from sales. Also, software producers tend to undertake efforts to make programs obscure and difficult to transfer from one machine to another, in an effort to discourage unauthorized users.[1] Programs which are poorly written in order to prevent others from using them also make it much more difficult for the firm who developed the programs to maintain and modify them.

All of these effects (the lack of public disclosure and the obscure programming style) will tend to reduce social welfare. For these reasons we do not believe that reliance on trade secrecy in the field of computer software is a desirable state of affairs. However, the advantages of a copyright system, or of public support for the development of computer software, must be weighed against the administrative costs of these systems, which may be higher than for a system of trade secrets.

Let us add here a few remarks on the relative merits of a patent vs. a copyright system. One may argue that what is essential about a computer program is the idea behind it, not the specific form it assumes, as is the case with literary works, and that therefore patenting of computer algorithms is more appropriate than a copyright. But this could endanger creativity. A law that grants property rights to producers of computer software should be formulated so as to prohibit any deliberate copying, which is an unambiguous violation. But it should not prohibit anyone from developing new programs, based only on ideas generally available in the scientific literature, and on a programmer's own original ideas. Such a policy would unduly hamper progress. For example, methods for the solution of linear programs are described in many standard textbooks. It should not be against the law to translate any such algorithm into a program, only because someone else has done this before. Only if it is obvious that the program has been copied from a copyrighted code should this be prohibited. People may well be expected to be honest, and it is perfectly reasonable to request that before anyone copies someone else's code, he assure himself whether he is entitled to do so, or may have to pay a fee to the copyright holder. But it would be unreasonable to expect from a creative person that before he develops a new computer program he should check whether anybody in the entire nation has ever written a similar program, and that he might therefore be liable of (unwitting) imitation. Such a provision of the law would be a paralyzing fetter on creative minds. For this reason, and because of the difficulty to prove the novelty for a patentable item, we consider a copyright law to be preferable to a patent law for computer software.

---

[1] Some ways in which this is done were discussed in Chapter III-A.

Is there any empirical information on how the amount of consumers' and producers' surplus from computer software obtained in reality compares with what is theoretically feasible? We have seen that in order to achieve the theoretical maximum of social welfare, all those software projects which yield benefits to society higher than the costs of production ought to be undertaken. Our analysis showed that since the revenue to a producer is only some fraction of the benefits to society, he will produce less software than is optimal for society. Some of the projects near the borderline of profitability will have a benefit to cost ratio greater than unity for society, but a revenue to cost ratio of less than one for the firm, and will therefore not be undertaken.[1] But from our theoretical analysis we would expect that, at least in the case of protection, the producer's revenue is a reasonably large fraction of the social benefits,[2] and that therefore all those projects whose social benefits substantially exceed their costs will be undertaken, even by private firms. But an empirical sutdy recently performed by Mathematica, Inc. shows evidence that this is not true.[3]

NASA developed a computer program, called NASTRAN, for the stability of flying rockets and their resistance to stress and vibrations of various frequencies. The same program can also be used to study the structural stability of airplanes and large buildings, for example a building's ability to withstand shocks from earthquakes of various amplitudes and frequencies. NASA made this program available for free to any individual or firm who wanted to use it. The costs to NASA for developing this program (in constant 1976 dollars, at an annual discount rate of 5 percent) were equal to 12 million dollars. The estimated economic benefits to all the firms who used it were estimated at $684 million. This corresponds to a benefit to cost ratio of 58:1. Doubtless, sooner or later someone would have undertaken to develop a similar project, given its great economic advantages. For the calculation of the above benefit to cost ratio, it was assumed that NASA's contribution only accelerated this development by 4 years. If it is assumed that it would have taken 10 years until an equivalent program had become available, then the benefit to cost ratio increases to a staggering 145:1. A few large firms, such as Ford Motor Company and Computer Sciences Corporation estimated their annual benefits from using the program to be comparable to its production costs of $12 million.

---

[1] On the order of about 50 percent, depending on the assumptions about the shape of the demand curve made.

[2] This is true, to various extents, both for the case of trade secrecy and of copyright protection.

[3] See "Quantifying the Benefits to the National Economy from Secondary Applications of NASA Technology," prepared for NASA by Mathematica, Inc., June 30, 1975.

The question arises why did no private firm spend the resources to develop such a program, given its high benefits? There are a number of possible reasons. First of all, it is impossible for a producer of software to realize all the benefits himself, through sales, because of the public goods nature of computer software. In addition, it is difficult to market a software package and sell it to all potential users, even to those who could derive benefits exceeding the sales price, because it is difficult to convince the management of a firm of the benefits of something new and unfamiliar. NASA had an advantage in this respect insofar as its prestigious name will be taken as a guarantee for quality by many users. Furthermore, if something is free, many potential users will be willing to take the risk of spending some time to try it out, and as they get to know it, they begin to appreciate it. This still leaves unexplained the fact that some large companies, who by themselves derived higher benefits from the program than its production costs, still did not invest the resources to develop something equivalent. A possible explanation is that top management in those companies is overly conservative and only willing to invest in something which has immediate tangible benefits, but they invest too little in more fundamental research and development.

Another area, where most of the fundamental research and innovation is performed at universities or other government supported research institutions, is agriculture. Most of the new seed strains, insecticides, or fertilizers have been developed at public non-profit institutions, not by farmers.[1] Not even large agribusiness enterprises invest an appreciable amount in innovation. Observations of this sort would indicate that governments do have to assume an important role in supporting basic research and innovation, including probably certain types of computer software.

One form, among others, in which the government could support the development of new software is through open competitions, with rewards for the most elegant and practical solutions. An institution similar to Research Applied to National Needs (RANN) could solicit proposals for important problems that are waiting for a solution. It could then send out and publish these lists, with a deadline for submitting porposed solutions in the form of computer programs, written in a standard language, which do the required job. The authors of the best solutions could be rewarded, and the results published and made available to the general public without payment. Many programmers have great talent in finding an ingenious solution to a problem they are given, but lack the ability to market the products of their work. They normally end up doing repetitive types of work with very narrow, specialized applications, for some company. But if they were given an opportunity to solve problems of a more general nature with wider applications, and could make their programs available to anyone who can use them, they would be able to make a much greater contribution to public welfare.

---

[1] According to a report in the New York Times of April 14, 1977, a research team at Michigan State University has isolated a substance called triacontanol from alfalfa, which in the tiny amounts of 4 milligrams per acre can increase the yield of many food crops by 10 to 40 percent. The enormous cost savings that this discovery can bring to people all over the world are hard to imagine.

How important is it to have effective institutions to encourage the production of new computer software? The costs of computer hardware, per arithmetic operation, have declined steadily at a rate of 55 percent per year over the last 20 years,[1] and there are good indications that this trend will continue for some time to come.[2] With this trend, computers will begin to play an increasing role in our daily lives. Perhaps the most important role they can play lies in the automation of many production processes, which can free people to devote their time to more creative activities than monotonous, repetitive tasks. Computers will play an increasingly important role in the area of storage and transmission of information, potentially replacing libraries and the publishing industry as we know it today. Other areas where automation has wideranging applications are banking operations, transportation (with automatically guided public transportation there is no need for drivers), in communication (human telephone operators will no longer be needed), even in agriculture where plans exist for robots to take over the decreasing number of tasks still left to people.

Automation is seen by many people today as a threat to their jobs and income, and is resisted on those grounds. This is understandable, because the potential benefits from automation have not always been shared equally in the past, benefitting employers at the expense of employees who were replaced by machines. But potentially automation could be used to increase the income and reduce the workload of everybody, not just a priviledged few. We anticipate that in time workable solutions to these problems of equity will be found.

---

[1] This figure is taken from Economics of Academic Libraries by W.J. Baumol and M. Marcus (American Council on Education, Washington, DC: 1973). Data used covered the years 1951-1967, but a similar trend has persisted since then.

[2] Many advances in the miniaturization of large scale integrated circuits have already been made, but not yet fully exploited in large computers available today. A price collapse, such as the one that took place for pocket calculators over the last few years, is likely to follow for more powerful computers as well.

With an increase in automation, the computer and software industry will not be limited to a few percent of GNP, but will become a major, perhaps even the major industry. Then it will be extremely important that the economy can develop, with a sufficient amount of easily usable software available, rather than still being plagued by incompatibilities between different types of hardware and different conventions on computer software[1] and by a sprawling proliferation of programming languages.[2] A failure to solve these problems satisfactorily today may be a major retarding factor for economic development in the future.

---

[1] Such incompatibilities make the life of programmers unnecessarily difficult and frustrating, and cause an enormous waste in human talent.

[2] A survey conducted by the Office of Naval Research revealed that the Pentagon is using more than 1000 different programming languages!

APPENDIX IV-1

Derivation of the Optimal Protection Period for a Linear Demand Curve

We shall calculate here what period of copyright protection maximizes social welfare, given demand for software can be represented by a linear demand curve (as shown in Figure 1a). To do this, we will compute the relative efficiency of any given period of protection, where the relative efficiency is defined as the ratio between the amount of social welfare forthcoming under a given duration of protection, and the maximum amount of welfare that is possible in principle with ideal central planning. Then we will determine the optimal protection period T* which maximizes the relative efficiency.

In order to calculate the amount of welfare generated by any given protection period, T, we will first derive the maximum revenue R(T) that the producer of a unit of software can generate through monopoly pricing during the period T. Then, by setting the incremental cost of the last unit of software produced equal to revenue per unit, we can calculate what amount X(T) of software the firm will produce. Next we will compute the discounted stream of consumers' and producer's surplus, S(T), forthcoming before and after the expiration of the protection period from one unit of software. By multiplying the amount of software produced, X(T), by the surplus generated per unit, S(T), and subtracting total production costs, C(T), we obtain the amount of welfare, W(T), resulting from a protection period of length T.

In order to calculate the maximum amount of social welfare feasible under the given cost and demand structure, W*, we assume that the amount of software produced, X^*, is determined in such a way that the incremental cost of the last unit is equal to the social benefit generated per unit. It is assumed that software will be made available for free to all potential users, so that the consumers' surplus is equal to the entire area under the demand curve (e.g., area OQ_0P_0 in Figure 4). Again, the maximum possible amount of social welfare to be generated, W^*, is equal to

$$W^* = X^*S^* - C^*,$$

where X^* is the amount of software produced, S^* is the surplus generated per unit, and C^* is the total cost of producing X^* units of software.

For guide reference, we list here all the essential equations of the model, using the same numbering as in section B of this chapter:

(1a) $$p^{(o)} = \begin{cases} p_0\,(1-q/q_0 & \text{if } q \leq q_0 \\ 0 & \text{if } q > q_0 \end{cases}$$

linear demand curve

(2) $$p^{(t)}(q) = p^{(o)}(q)e^{-t/t_o}$$

exponential decay of the value of a unit of software

(3) $$MC(x) = c\cdot x^v$$

incremental cost of the x^{th} unit of software produced. The exponent v indicates how rapidly marginal costs increase with increasing output.

(4) $$p^{(t)}(q) = 0 \quad \text{for } t > T$$

expiration of the monopoly after T time units.

(5) $\frac{d}{dp}(p \cdot q) = 0$

for the monopoly price $p = p^*$

(6) $R(T) = \int_0^T q^* p^* e^{-t/t_o} e^{-rt} dt$

monopoly revenue during protection period T, given the discount rate is r and the average useful lifetime of a unit of computer software it t_o.

(7) $c \cdot X(T)^v = R(T)$

marginal revenue equals marginal cost

(8) $W(t) = \int_0^{q^*} p^{(t)} dq$ for $t \leq T$

consumers' and producer's surplus generated per time unit, per unit of software, during the protection period.

(8') $W'(t) = \int_0^{\infty} p^{(t)} dq$ for $t > T$

consumers' and producer's surplus generated per time unit, per unit of software after expiration of the protection period.

(9) $S(T) = \int_0^T W(t) e^{-rt} dt + W'(t) e^{-rt} dt$

discounted stream of consumers' and producer's surplus generated per unit of software, during and after the period of protection

(10) $C(T) = \int_0^{X(T)} MC(x) dx$

total cost of producing X(T) units of software

(11) $W(T) = X(T)S(T) - C(T)$

amount of social welfare generated through a protection period of length T (by a single firm).

We begin by calculating the amount of software a producer will supply, and the price he will charge per unit, as long as he is granted a monopoly. If the demand curve is given by

$$q = q_o(1-p/p_o),$$

the linear shape shown in Figure 1a, then the revenue per time unit that the monopolist can earn, if he charges price p, is

$$r(p) = p \cdot q = pq_o(1-p/p_o).$$

By taking the derivative of r(p) with respect to p and setting it equal to zero, it is easy to see that the price which maximizes the monopolist's revenue in this case is equal to $p^* = p_o/2$. This applies at time $t = 0$. After a time span t has elapsed, all the prices consumers are willing to pay decline by a factor $f(t) = e^{-t/t_o}$. The monopoly price also declines to

$$p^{*(t)} = p^* e^{-t/t_o} = \frac{p_o}{2} e^{-t/t_o} \quad ,$$

but the quantity q^* sold by the monopolist remains unchanged.

The discounted stream of revenue a monopolist can earn during the life-time of copyright protection is equal to

$$\begin{aligned} R(T) = & \int_o^T q^* p^* e^{-t/t_o} e^{-rt} dt \\ = & \frac{q_o p_o}{4} \int_o^T e^{-(r+1/t_o)t} dt . \end{aligned}$$

Using the notation

$$s = r + 1/t_o ,$$

we obtain

$$R(T) = \frac{q_o p_o}{4} \int_o^T e^{-st} dt$$

$$= \frac{q_o p_o}{4} \cdot \frac{1}{s} \; (1-e^{-sT}).$$

To find the amount being produced by the monopolist, X(T), we set marginal revenue equal to marginal cost:

$$c \cdot X(T)^v = R(T) \qquad \text{or} \qquad X(T) = \left(\frac{1}{c} R(T)\right)^{1/v}$$

At time $t = 0$, the producer's surplus (corresponding to area OQRP in Figure 4) will be equal to $p^* q^* = p_o q_o/4$. The consumers' surplus (the small triangle corresponding to area PRP_0 in Figure 4) will be equal to $p_o q_o/8$. Thus, at time $t = 0$, the consumers' and producer's surplus generated by one unit of software is equal to $(3/8)p_o q_o$. At time t during the protection period, the contribution to the surplus is

$$W(t) = (3/8)p_o q_o \; e^{-t/t_o} \; .$$

After protection has elapsed, the contribution to the consumers' surplus (area OQ_0P_0 in Figure 8) is equal to

$$W'(t) = (1/2)p_o q_o \; e^{-t/t_o} \; .$$

The discounted stream of consumers' and producer's surplus generated by one unit of software is equal to

$$S(T) = \int_0^T W(t)e^{-rt}dt + \int_T^\infty W'(t)e^{-rt}dt$$

$$= \int_0^T (3/8)p_oq_oe^{-(r+1/t_o)t}dt+\int_T^\infty (1/2)p_oq_oe^{-(r+1/t_o)t}dt$$

$$= (3/8)p_oq_o \cdot \frac{1}{s}(1-e^{-st})+(1/2)p_oq_o \cdot \frac{1}{s} e^{-st}$$

$$= (1/2)p_oq_o \cdot \frac{1}{s} - (1/8)p_oq_o \frac{1}{s}(1-e^{-st}) ,$$

where $s = r + 1/t_o$.

The total cost of producing $X(T)$ units of software is

$$C(T) = \int_0^{X(T)} c \cdot x^v dx = c \cdot X(T)^{v+1}/(v+1) .$$

Finally, the amount of welfare generated through a protection period of length T is equal to

$$W(T) = X(T)S(T) - C(T).$$

Using the notation

$$z = 1 - e^{-st} ,$$

and substituting all the expressions we obtain

$$W(T) = \left(\frac{p_oq_o}{4sc} z \right)^{1/v} \cdot \frac{p_oq_o}{2s} (1-z/4) - \frac{c}{v+1} \left(\frac{p_oq_o}{4sc} z \right)^{(v+1)/v} .$$

We now go on to calculate the maximum amount of social welfare that can be achieved if the government purchases the socially optimal amount of software from the firm under consideration[1], and makes it available without payment to the entire society. In that case, we have the set of equations

---

[1] The way this would happen in practice is that the government sets a price for software it purchases, which is equal to the estimated marginal social benefit, and then lets the firm produce as much as it wants at the given price.

(17) $$S^* = \int_0^\infty W'(t)e^{-rt}dt,$$

(18) $$X^* = (\frac{1}{c}S^*)^{1/v}$$

(19) $$C^* = \frac{c}{v+1}(X^*)^{v+1}, \quad \text{and}$$

(20) $$W^* = X^*S^* - C^* .$$

We find

$$S^* = \int_0^\infty (1/2)p_o q_o e^{-(r+1/t_o)t}dt = \frac{p_o q_o}{2} \cdot \frac{1}{s} .$$

The optimal amount of software to be produced is

$$X^* = (\frac{1}{c}S^*)^{1/v} = (\frac{p_o q_o}{2sc})^{1/v} .$$

The cost of production is equal to

$$C^* = \frac{c}{v+1}(\frac{p_o q_o}{2sc})^{(v+1)/v} .$$

The maximum feasible amount of welfare is therefore

$$W^* = X^*S^* - C^* = (\frac{p_o q_o}{2sc})^{1/v} \cdot \frac{p_o q_o}{2s} - \frac{c}{v+1}(\frac{p_o q_o}{2sc})^{(v+1)/v}$$

$$= (\frac{1}{c})^{1/v}(\frac{p_o q_o}{2s})^{(v+1)/v}(1 - \frac{1}{v+1}) .$$

We are now able to compute a coefficient of relative efficiency, $E(T)$, for a protection period of length T:

$$E(T) = W(T)/W^*.$$

After some algebraic manipulations, one will find

$$(15) \qquad E(T) = \frac{1}{v} \left(\frac{z}{2}\right)^{1/v} \left(v+1 - \frac{v+3}{4} z\right),$$

where $z = 1 - e^{-sT}$.

The optimal protection period T^* is the one which maximizes the relative efficiency $E(T)$. It corresponds to the optimal value z^* for z. Taking the derivative of $E(T)$ with respect to z, and setting it equal to zero, we obtain

$$\frac{d}{dz} E(T) = \frac{1}{v} \left(\frac{z}{2}\right)^{1/v-1} \left(\frac{1}{2}\right) \cdot \frac{1}{v} \left(v+1 - \frac{v+3}{4} z\right) + \left(\frac{z}{2}\right)^{1/v} \cdot \frac{1}{v}\left(- \frac{v+3}{4}\right) = 0.$$

By simplifying and solving for $z = z^*$ we find

$$z^* = \frac{4}{v+3} \quad .$$

Substituting back

$$z^* = 1 - e^{-sT^*},$$

we finally get the formula

$$(13) \qquad T^* = \frac{1}{s} \ell n \frac{v+3}{v-1} = \frac{1}{r+1/t_o} \ell n \frac{v+3}{v-1} \quad .$$

APPENDIX IV-2

Derivation of the Optimal Protection Period for an Example of a Nonlinear Demand Curve.

In this Appendix we will calculate the relative efficiency of a period of copyright protection T and the optimal duration T* for the case of the nonlinear demand curve shown in Figure 1b. All the steps are essentially the same as in Appendix IV-1, except that the demand equation (1a) is replaced by the equation

$$(1b) \qquad p^{(o)}(q) = k(q+q_o)^{-u}, \qquad \text{where } u > 1.$$

We will therefore limit ourselves to indicating the major steps in the calculation, and refer for their motivation to Appendix IV-1.

To maximize his revenue from each unit of software produced, the monopolist, at time t = 0, will sell q* copies of each unit of software he has produced, where for q = q* we have

$$(5) \qquad \frac{d}{dq}\left(p^{(o)}(q) \cdot q\right) = 0 = \frac{d}{dq}\left(k(q+q_o)^{-u}q\right)$$

$$= kq(-u)(q+q_o)^{-u-1} + k(q+q_o)^{-u} .$$

This yields the monopoly quantity and price

$$q^* = \frac{q_o}{u-1} \quad \text{and} \quad p^* = k\left(\frac{u}{u-1}q_o\right)^{-u} .$$

The revenue earned by the monopolist is equal to

$$(6) \qquad R(T) = \int_0^T p^*q^*e^{-(r+1/t_o)t} = p^*q^*\int_o^T e^{-sT}dt = \frac{kq_o}{u-1}\left(\frac{uq_o}{u-1}\right)^{-u} \cdot \frac{1}{s}(1-e^{-sT}).$$

We will again use the abbreviation

$$z = 1-e^{-sT} \quad .$$

By setting revenue per unit equal to marginal cost, we obtain for the quantity produced

$$(7) \qquad X(T) = \left(\frac{1}{c} R(T)\right)^{1/v} \quad .$$

During the protection period, consumers' and producer's surplus generated by one unit of software is equal to

$$(8) \qquad W(t) = e^{-t/t_o} \int_o^{q^*} p dq = e^{-t/t_o}\, k \int_o^{q^*} (q+q_o)^{-u}\, dq.$$

Solving the integral and inserting $q^* = q_o/(u-1)$ we obtain

$$W(t) = e^{-t/t_o} \cdot \frac{k}{u-1} \left(q_o{}^{1-u} - \left(\frac{uq_o}{u-1}\right)^{1-u}\right) \quad .$$

Similarly we find for the surplus after the expiration of the protection period

$$(8') \qquad W'(t) = e^{-t/t_o} \cdot \frac{k}{u-1}\, q_o{}^{1-u} \quad .$$

The discounted stream of consumers' and producer's surplus generated by a unit of software is

$$(9) \qquad S(T) = \int_o^T W(t) e^{-rt} dt + \int_T^\infty W'(t) e^{-rt} dt = \frac{k}{u-1} \cdot \frac{1}{s} \left(q_o{}^{1-u} - \left(\frac{uq_o}{u-1}\right)^{1-u} \cdot z\right)$$

with $z = 1 - {}^{e-st}.$

Total cost is given by

$$(10) \qquad C(T) = \frac{c}{v+1} X(T)^{v+1} \quad .$$

Combining the expressions (6), (7), (9) and (10) yields

$$(11) \qquad W(T) = X(T)S(T) - C(T),$$

$$= \left[\frac{kq_o}{sc(u-1)} \left(\frac{uq_o}{u-1}\right)^{-u} z\right]^{1/v} \left[\frac{k}{s(u-1)} \left(q_o^{1-u} - \left(\frac{uq_o}{u-1}\right)^{1-u} z\right)\right]$$

$$- \frac{c}{v+1} \left[\frac{kq_o}{sc(u-1)} \left(\frac{uq_o}{u-1}\right)^{-u} z\right]^{(v+1)/v} .$$

For the maximum social welfare feasible under public financing we evaluate formulas (17) to (20) and obtain

$$(17) \qquad S^* = \int_0^\infty W'(t) e^{-rt} dt = \frac{k}{s(u-1)} q_o^{1-u}$$

$$(18) \qquad X^* = \left(\frac{1}{c} S^*\right)^{1/v} = \left(\frac{k}{sc(u-1)} q_o^{1-u}\right)^{1/v}$$

$$(19) \qquad C^* = \frac{c}{v+1} (X^*)^{v+1}$$

$$(20) \qquad W^* = X^*S^* - C^* = \frac{v}{v+1} c\left(\frac{k}{sc(u-1)} q_o^{1-u}\right)^{(v+1)/v} .$$

We can now compute the relative efficiency $E(T)$ of a given protection period T as

$$(16) \qquad E(T) = W(T)/W^* = \frac{v+1}{v} (yz)^{1/v} \left(1 - \frac{u}{u-1} yz\right) - \frac{1}{v} (yz)^{(v+1)/v}$$

where z stands for $1 - e^{-sT}$ and y for $(u/(u-1))^{-u}$.

In order to find the optimal protection period T^*, we take the derivative of the coefficient of relative efficiency, $E(T)$, with respect to z and set it equal to zero. This yields

$$z^* = \left(\frac{u}{u-1}\right)^u \cdot \frac{u-1}{uv+2u-1} \quad \text{and} \quad T^* = -\frac{1}{s} \ln(1 - z^*).$$

By replacing s by $r + 1/t_o$ we obtain equation (14) in Section IV-B.